

The Uncanny Valley Hypothesis: behavioural, eye-movement, and functional MRI findings

Thesis
presented to the Faculty of Arts
of
the University of Zurich
for the degree of Doctor of Philosophy

by
Marcus Cheetham
of Bowdon / England

Accepted in the fall semester 2011 on the
recommendation of Prof. Dr. Lutz Jäncke and Prof. Dr. Martin Meyer

Zurich, 2014

CONTENTS

Acknowledgements	4
Summary	5
Zusammenfassung	9
1. Introduction.....	13
2. Theoretical background.....	14
3. Study A. The human likeness dimension of the "uncanny valley hypothesis": behavioural and functional MRI findings.....	23
3.1. Abstract	23
3.2. Introduction	24
3.3. Material and methods.....	29
3.4. Experiment 1: Forced choice classification.....	30
3.5. Experiment 2: Perceptual discrimination.....	31
3.6. Experiment 3: Target monitoring and fMRI	34
3.7. Results.....	37
3.8. Discussion.....	46
3.9. References.....	53
4. Study B. Category processing and the human likeness dimension of the uncanny valley hypothesis: eye-tracking data	60
4.1. Abstract	60
4.2. Introduction	60
4.3. Experiment 4. Category processing and eye-movement.....	65
4.4. Materials and Methods.....	65
4.5. Results.....	69
4.6. Discussion.....	76
4.7. References.....	82
5. Study C. Perceptual discrimination difficulty and familiarity in the Uncanny Valley.....	89
5.1. Abstract	89
5.2. Introduction	89
5.3. Experiment 5: Perceptual discrimination and forced choice categorization.	93
5.3.1 Materials and Methods.....	93
5.3.2. Results	97
5.3.3. Discussion.....	101
5.4. Experiment 6. Perceptual discrimination and familiarity.	103

5.4.1. Materials and Methods	103
5.4.2. Results	104
5.4.3. Discussion.....	110
5.5. Experiment 7. Perceptual discrimination and differential processing bias.....	112
5.5.1. Materials and Methods	112
5.5.2. Results	112
5.5.3. Discussion.....	115
5.6. General discussion	116
5.7. References.....	123
6. General discussion.....	128
7. References	135
8. Curriculum Vitae	141

Acknowledgements

I warmly thank Prof. Lutz Jäncke for his attentive and enthusiastic support in pursuing a line of research that has hitherto received only scant attention in cognitive neuropsychology. I thank him and the teaching and research staff of the Department of Neuropsychology, University of Zurich, for sharing with me a scientifically rich working environment. Of the many colleagues and co-workers I would like to thank for this, my particular appreciation goes to my co-workers Dr. Andreas Pedroni, Dr. Kai Lutz, and Dr. Jürgen Hänggi for their instrumental support in brain research methodology, and to Pascal Suter for his keen interest and attention to detail. My appreciation goes to Prof. Martin Meyer for his thoughtful suggestions and to Heinz Blaser, Dr. Valentine Marcar, and Dr. Karin Nadig for their invaluable assistance. I express my special gratitude to my family for their unfailing patience, support and encouragement.

Summary

The subjective perception and experience of characters in media such as novels, films, and virtual reality-based games and environments has long engaged interest (Tellegen and Atkinson, 1974; Lombard and Ditton, 1997; Maccoby and Wilson, 1957; Horton and Wohl, 1956; Balázs, 1938). Especially in visually-based media, advances in sophisticated technology and computer animation, the desire to meet the challenge of high realism (MacGillivray, 2007; Maddock et al., 2005), and efforts to improve the usability of virtual reality (Geven et al., 2006; Power et al., 2002) have shifted attention to the affect on subjective experience of high realism in the anthropomorphic design of humanlike characters. This attention has been heightened by the general concern that greater realism does not necessarily predict a better reception of the character by the observer (Geller, 2008). One particular expression of this concern is summarized in the so-called *Uncanny Valley Hypothesis (UVH)* (Mori, 1970). The UVH suggests that particular degrees of anthropomorphic realism in the design of characters and objects (e.g., robots, prosthetic hand) elicit negative affective experience akin to various forms of avoidant behaviour, failing therefore to engage the user in the way that the character's or object's design intended. Though Mori did not subject his hypothesis to empirical examination, the UVH has influenced animators, video game designers, and roboticists in the design of anthropomorphic characters and robots and lead to a relatively new field of research.

This new field is still in a formative stage in which many potential aspects of the impact on subjective experience of anthropomorphic characters and objects have yet to be investigated. The present thesis focuses on developing a better understanding of the *dimension of human likeness (DHL)* of the UVH. This focus was motivated by the fact that the UVH seeks to understand positively and negatively valenced affective experience as a function of human likeness along the DHL. While the examination of affective experience has received much attention, the way in which human likeness along the DHL is actually perceived has received no attention. Given the absence of any research on the DHL, the general inconsistencies in uncanny-related findings to date, and the likelihood that the UVH' conceptualization of human likeness does not reflect the way in which human likeness is perceived and

represented cognitively, three studies comprising in total seven experiments were conducted to develop more insight into the UVH. These studies used morph continua based on human and highly realistic computer-generated (avatar) faces to represent the DHL.

Study A entitled “The human likeness dimension of the “uncanny valley hypothesis”: behavioural and functional MRI findings” had two aims. The first aim was to determine whether the DHL could be defined in terms of effects of *categorical perception (CP)*. In contrast to the UVH’ definition of the DHL as a linear perceptual dimension, the two behavioural experiments show that *perceptual discrimination (PD)* performance between faces along the DHL is enhanced in that region of the DHL at which the UVH anticipates that it should be attenuated. This is critical for the UVH because it assumes that attenuated PD (i.e., increased perceptual discrimination difficulty) evokes negative affective experience. These findings suggest that this prediction might be wrong (see Study C). The second aim was to examine the impact of processing category change along the DHL. The expectation was that categorization experience with novel nonhuman but highly humanlike faces would evoke a different pattern of neural activity than that evoked by human faces (for which there is everyday perceptual and categorization expertise). The event-related functional MRI study (the third experiment) confirmed this, showing that a change in category between (otherwise highly similar) sequentially presented faces evoked modulations of neural activity in brain regions associated with category learning and category uncertainty. Importantly, the pattern of regions was entirely different depending on whether the target of processing was an avatar or human face. This finding is important for the UVH and the DHL because it suggests that extremely similar nonhuman and human category faces represent different categorization problems that require dissimilar processes or strategies to resolve them.

Study B “Category processing and the human likeness dimension of the uncanny valley hypothesis: eye-tracking data” followed up on this suggestion. It examined whether there is a difference in eye sampling behaviour when extracting perceptual information from faces of the DHL to determine the human versus nonhuman category membership of the faces. Differences were compared between categorically

unambiguous human and nonhuman faces and categorically ambiguous faces using the dependent measures *total number of fixations* and *dwell time of fixations* to facial features. Both the data for fixation number and dwell time confirmed the relative importance of the face feature hierarchy (eyes, nose, and mouth) across the DHL, but there were no further findings for fixation number. Categorization ambiguity did influence the relative amount of dwell time spent extracting visual information from the eye, nose, and mouth before making a category decision, but this effect applied only for unambiguous nonhuman compared with ambiguous faces. The behavioural data showed that category assignment of nonhuman faces is much faster than that of human faces, meaning that there is a categorization advantage for nonhuman faces. Considered together, the data show that categorically unambiguous nonhuman faces are processed differently than ambiguous or human faces. This difference is consistent with the idea that category decision making is generally influenced by a cognitive strategy that involves preferential detection of nonhuman-specifying rather than human-specifying perceptual information along the HDL. What this information might be was investigated in the final experiment of Study C.

Study C entitled “Perceptual discrimination difficulty and familiarity in the Uncanny Valley” was based on consideration of the findings and conclusions of the experiments in Studies A and B. The critical feature of the UVH is the prediction that perceptual discrimination difficulty will evoke negatively valenced (i.e., uncanny) experience at that point along the DHL at which categorization is most ambiguous. In the first experiment of Study C, a perceptual discrimination task was used to delineate the profile of PD difficulty along the DHL. The findings clearly reject the implicit assumption underlying the UVH' prediction, showing that PD difficulty is in fact attenuated for categorically ambiguous faces and greatest for unambiguous human faces. In addition, and after controlling for physical differences, the data show a clear asymmetry in PD difficulty such that nonhuman faces (and ambiguous faces) are easier to discriminate than human faces. The second experiment replicated the findings of the first. Importantly, the second experiment tested the UVH' predicted direction of the relationship between PD difficulty and affective experience (in the sense of the UVH' *familiarity* dimension). In contrast to the UVH, greater PD difficulty was associated with more positively rather than negatively valenced affective experience. Notably, this effect was actually strongest for those faces along the DHL

that were categorically most ambiguous. A third experiment examined a possible explanation for the asymmetry in PD difficulty along the DHL. This experiment found no evidence to suggest that the asymmetry in PD along the DHL is attributable to a differential processing strategy (i.e., processing avatars at a category and human faces at an individual level). But the data indicated that PD along the DHL generally relies on the coding of human likeness-specifying information relating to facial features such as the eyes, nose, and mouth and other features such as skin tone. In conclusion, the data strongly challenge the UVH by showing that the UVH' assumed distribution of PD difficulty along the DHL and the predicted relationship between this and feelings of familiarity is incorrect.

Zusammenfassung

Die subjektive Wahrnehmung und das Erleben von fiktiven Charakteren in verschiedenen Medien wie Romane, Filme, Videogames und Virtual Reality Environments erweckt seit langem intensives Interesse (Tellegen und Atkinson, 1974; Lombard and Ditton, 1997; Maccoby und Wilson, 1957; Horton und Wohl 1956, Balázs, 1938). Vor allem im Bereich der visuell-basierten Medien haben Fortschritte in der ausgefeilten Technik und Computeranimation, kombiniert mit dem Ziel noch höheren Realismus zu erreichen (MacGillivray, 2007; Maddock et al, 2005) und die Benutzerfreundlichkeit der virtuellen Realität zu verbessern (Geven et al, 2006; Power et al, 2002), die Aufmerksamkeit von Forschern und Designern auf die Frage nach der Auswirkung sehr realistischer menschenähnlicher Charakteren auf subjektive Wahrnehmung und Erleben gelenkt.

Dieses Interesse hat v.a. mit der Annahme zu tun, dass erhöhter Realismus im anthropomorphischen Design nicht unbedingt zu einer besseren Akzeptanz des Charakters führt (Geller, 2008). Diese Annahme wird wohl am besten in der sogenannten *Uncanny Valley Hypothese (UVH)* (Mori, 1970) zum Ausdruck gebracht. Die UVH legt nahe, dass insbesondere anthropomorphisch sehr realistische Charaktere und Objekte (zB. Roboter, Hand-Prothese) einen negativen Affekt auslösen können. Obwohl Mori seine Hypothese nicht empirisch geprüft hat, ist der Einfluss der UVH auf Animatoren, Videospiel-Designer und Roboteringenieure dennoch groß (zB. Minato et al., 2006). Dieser Einfluss hat auch zu einem relativ neuen Bereich der Forschung geführt.

Dieser Forschungsbereich befindet sich in einer frühen Phase, in der noch viele Aspekte der Auswirkung von anthropomorphen Charakteren und Objekten auf das subjektive Erleben unbekannt sind. Die Studien in dieser Abhandlung fokussieren auf die Entwicklung eines besseren Verständnisses der Dimension der Menschenähnlichkeit (*DM*) der UVH, d.h. wie Stimuli entlang der DM wahrgenommen werden, und wie diese Wahrnehmung im Zusammenhang zu erlebtem Affekt steht. Die UHV versteht das Auftreten von negativem und positivem Affekt in Abhängigkeit des Ausmasses an *Menschenähnlichkeit* entlang der DM. Bisherige Uncanny Studien haben sich auf die Affektdimension der UVH fokussiert, aber nichts ist darüber

bekannt, wie Menschenähnlichkeit entlang der DM wahrgenommen wird. Angesichts der inkonsistenten Befunde der bisherigen Uncanny Studien über den Uncanny-Effekt, der fehlenden Kenntnisse über die DM, und der Wahrscheinlichkeit, dass die UVH-Konzeption der DM, wie Menschenähnlichkeit wahrgenommen und kognitiv repräsentiert wird, falsch ist, wurden drei Studien (bestehend aus insgesamt sieben Experimenten) durchgeführt. Jede Studie nutzt Morphkontinuen, um die DM zu repräsentieren. Die Kontinuen wurden auf Basis von Menschenbildern und höchst realistischen computergenerierten (Avatar) Bildern entwickelt.

Studie A (Titel: "The human likeness dimension of the "uncanny valley hypothesis": behavioural and functional MRI findings") hatte zwei Ziele. Das erste Ziel war es, festzustellen, ob Effekte der kategorialen Wahrnehmung einen Einfluss auf die kognitive Repräsentation von Stimuli entlang der DM haben. Im Gegensatz zu der Definition der DM in der UVH, die die DM als eine lineare physikalische Dimension versteht, zeigten die beiden Verhaltensexperimente der Studie A, dass die perzeptuelle Diskriminierungsleistung entlang der DM sehr unterschiedlich (d.h. nicht linear) ist. Vor allem in dem Bereich der DM, in dem die UVH annimmt, dass die perzeptuelle Diskriminierung schwieriger sein soll, war diese Leistung deutlich besser. Dieser Befund ist entscheidend für die UVH, weil sie vorhersagt, dass gerade die erhöhte Schwierigkeit der perzeptuellen Diskriminierung negativen Affekt auslöst. Diese Ergebnisse legen nahe, dass diese Vorhersage der UVH falsch sein könnte. Das zweite Ziel war es, die Auswirkungen der kategorialen Verarbeitung entlang der DM auf neuronale Prozesse zu untersuchen. Die ereignis-korrelierte funktionelle MRI-Untersuchung bestätigt, dass die kategoriale Verarbeitung von Avatar und Menschengesichtern zur Modulation neuronaler Aktivität in Hirnregionen führt, die mit kategorialem Lernen und kategorialer Ambiguität assoziiert sind. Dabei zeigte sich, dass das Muster an Hirnregionen unterschiedlich war, je nachdem, ob der Zeilreiz bei der kategorialen Verarbeitung ein Avatar oder ein Mensch war. Das unterschiedliche Muster ist assoziiert mit unterschiedlicher Kategorisierungserfahrung (d.h. für neue höchstrealistische Avatars im Vergleich mit typischen Menschengesichtern). Diese Feststellung ist wichtig für den UHV, weil sie schlussfolgern lässt, dass die Verarbeitung von sehr realistischen Avatars und Menschen unterschiedliche Kategorisierungsprobleme darstellen, die durch unterschiedliche Prozesse oder kognitive Strategien gelöst werden.

Studie B ("Category processing and the human likeness dimension of the uncanny valley hypothesis: eye-tracking data") ging auf diese Schlussfolgerung ein und prüfte, ob es einen Unterschied bei kategorialen Wahrnehmungsentscheidungen in den Blickbewegungen gibt. Dabei dienen die Blickbewegungen dem Extrahieren von perzeptuellen Informationen zur Entscheidungsfindung. Dazu wurden die Variablen Fixationsdauer und Fixationshäufigkeit von bestimmten Gesichtsmerkmalen bis zum Zeitpunkt der Wahrnehmungsentscheidung erfasst und analysiert. Die Daten bestätigten die relative Bedeutung der Hierarchie der Gesichtsmerkmale (d.h. Augen, Nase, Mund in dieser Reihenfolge) entlang der gesamten DM. Die Daten zeigten aber auch, dass sich kategoriale Ambiguität auf die Fixationsdauer auf diese Gesichtsmerkmale auswirkte, aber nur im Vergleich zwischen eindeutigen Avatars und uneindeutigen Gesichtern. Die Verhaltensdaten zeigten einen Kategorisierungsvorteil für die nicht-menschlichen Gesichter, in dem die Kategorisierung von Avatars deutlich schneller war als die von Menschengesichtern. Zusammen berücksichtigt, sprechen die Befunde dafür, dass kategoriale Wahrnehmungsentscheidungen entlang der DM durch eine Klassifizierungsstrategie beeinflusst wird, in der perzeptuelle Information, die diagnostisch ist, für die Nichtmenschenkategorie präferentiell kodiert wird.

Studie C ("Perceptual discrimination difficulty and familiarity in the Uncanny Valley") basierte auf die Erkenntnisse und Schlussfolgerungen aus den Experimenten in Studien A und B. Die UVH macht die Vorhersage, dass Schwierigkeit bei der perzeptuellen Diskriminierung zwischen Stimuli der DM mit negativen Affekt (d.h. den Uncanny Effekt) einhergeht. Dabei soll der Uncanny Effekt an dem Punkt entlang der DHL auftreten, an dem die kategoriale Ambiguität zwischen Exemplare der Menschen- und der Nichtmenschenkategorie am grössten ist. Im ersten Experiment wurde das Profil der Schwierigkeit der *perzeptuellen Diskriminierung (PD)* entlang der DM erfasst. Die Ergebnisse zeigen deutlich, dass die implizite Annahme der UVH, dass PD Schwierigkeit an diesem Punkt am grössten ist, widerlegt werden kann. Im Gegensatz dazu, zeigten die Daten verbesserte PD für die kategorial uneindeutigen Gesichtern und grösste PD Schwierigkeit für die Menschenkategorie. Die Daten zeigen auch eine klare Asymmetrie in PD entlang der DM, in dem zwischen nicht-menschlichen Gesichtern generell leichter zu diskriminieren ist als

zwischen Menschengesichtern. Das zweite Experiment repliziert die Ergebnisse des ersten. Das zweite Experiment prüfte auch die in der UVH vorhergesagte Richtung der Beziehung zwischen PD Schwierigkeiten und erlebten Affekt (d.h. im Sinne der Familiaritätsdimension der UVH). Im Gegensatz zur Vorhersage der UVH ging zunehmende PD Schwierigkeit mit zunehmend positivem statt negativem Affekt einher. Dabei war dieser Effekt tatsächlich am stärksten für die kategorial uneindeutigen Gesichter der DM. Ein drittes Experiment untersuchte eine mögliche Erklärung für die Asymmetrie der PD Schwierigkeiten entlang der DHL. Dieses Experiment fand aber keinerlei Hinweise darauf, dass diese Asymmetrie auf eine *Differentielle Verarbeitungsstrategie* zurückzuführen ist (d.h., die Verarbeitung von Avatars auf der Kategorienebene und die menschlichen Gesichter auf der individuellen Ebene). Die Daten des dritten Experiments sprechen aber dafür, dass die perzeptuelle Diskriminierungsleistung entlang der DM generell auf die Kodierung von perzeptuellen Merkmalen (wie Augen, Nase und Mund) und andere Merkmale wie die Hauttextur (Farbe, Schattierung) beruht. Schlussfolgernd sagen die Daten aus, dass die in der UHV angenommene Verteilung der PD Schwierigkeiten entlang der DHL wie auch die vorhergesagte Richtung der Beziehung zwischen PD Schwierigkeiten und Affekt (d.h. Familiarität) sehr wahrscheinlich falsch ist.

1. Introduction

Inspired by informal observation of the emotionally aversive responses of individuals to their encounter with human-like robots, Mori (1970) put forward the Uncanny Valley Hypothesis (UVH). Though vague in its formulation, the UVH proposes a nonlinear relationship between an entity's degree of physical human likeness, defined along the UVH's *dimension of human likeness (DHL)* and the subjective experience of positively or negatively valenced feelings and cognitive evaluations in response to that entity. This experience is defined along the UVH's affective dimension of *familiarity* (see Figure 1). The key feature of this relationship is the prediction that individuals will experience a state of personal discomfort due to difficulty distinguishing a realistic humanlike entity (e.g., industrial robot, lifelike prosthetic hand, human corpse, mannequin) from its natural human equivalent. This state is described as including feelings of strangeness and the uncanny. This uncanny effect is reflected in a sharp negative peak or valley (thus dubbed the *uncanny valley*) in the slope of Mori's depiction of the relationship between human likeness and subjective experience.

The longstanding UVH has only recently evoked research interest (e.g., Hanson, 2005). The aim of this thesis is to develop further insight into the underlying notions of the UVH. It does so by introducing to uncanny-related research a focus on the DHL. The reason for this focus is that the UVH and research guided by it aims to understand affective experience of as a function of human likeness. In defining affective experience this way, a better understanding of the subjective perception of human likeness along the DHL is critical, especially in view of the general inconsistencies in findings to date in uncanny-related research and the UVH' conceptualisation of human likeness along the DHL. Given the absence of any research on the DHL, Chapter 2 of this thesis brings together the essential elements of uncanny-related research that form the theoretical background for the studies of this thesis and the issues in uncanny research that these studies seek to address. These studies are presented in Chapters 3, 4 and 5. Chapter 6 concludes the thesis with a summary of the results of these studies and a general discussion.

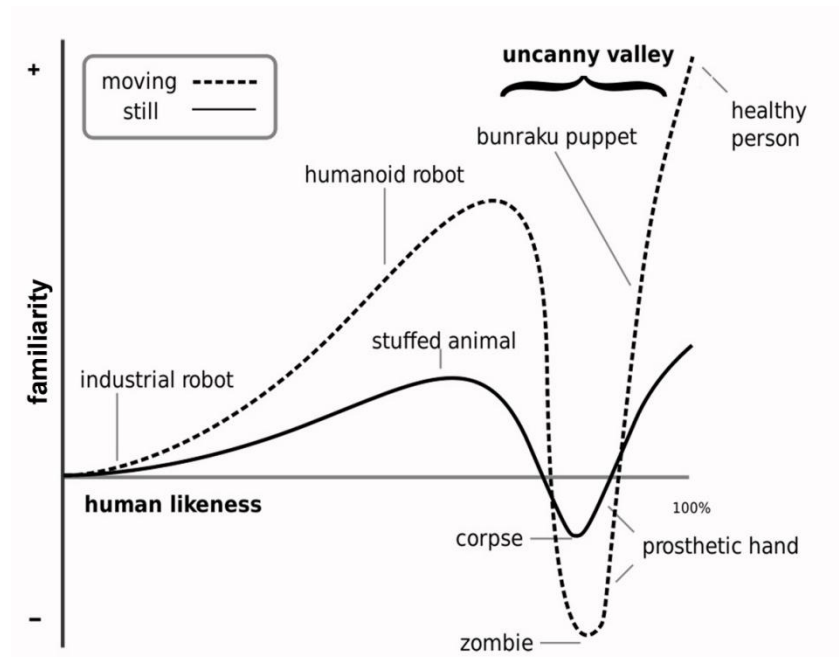


Figure 1. Illustration of the non-linear relationship between positively and negatively valenced subjective experience (i.e., the UVH' dimension of familiarity) and perceived human likeness (i.e., along the UVH' dimension of human likeness). The otherwise positive relationship shows a sharp negative peak (i.e., uncanny valley) at the level of realism at which fine-grained differences in appearance and behaviour between a highly realistic nonhuman entity and the human equivalent is suggested to elicit a sense of strangeness and personal discomfort, that is, an uncanny feeling (illustration from MacDorman, 2006).

2. Theoretical background

The progress of recent years in robotics and computer graphics in the realistic simulation of human appearance and behaviour in conjunction with ongoing uncertainty as to the real influence of enhanced anthropomorphism on subjective has spurred interest in the UVH in relation to the design of anthropomorphic robots and computer-generated characters experience (e.g., Ho & MacDorman, 2010; Minato et al., 2004; Woods et al., 2004; Walters et al., 2008; Fabri et al., 2004; Tinwell and Grimshaw 2009). This uncertainty has been fuelled in part by the intuitive appeal of the UVH and by anecdotal “evidence” indicating that high realism in computer animations of various feature films and featurettes (Ho and MacDorman 2010; MacDorman et al., 2009; Tinwell and Grimshaw 2009) and modern robots, for example, the Geminoid HI-1 and Repliee Q2 robots (Saygin et al., 2011; Becker-

Asano, 2010) can evoke a sense of disquiet. From this, a new domain of research has evolved that is very much in its formative stage.

This stage is characterised by a process toward understanding and specifying more clearly the general research problem stated in the UVH and the approach to its investigation. This process is apparent in the initial exploratory studies of the uncanny problem presented in the UVH (Hanson, 2005; MacDorman, 2006), subsequent doubt and criticism (Hanson, 2006), search for new supporting data (Seyama and Nagayama, 2007), formation of new ideas and opinions, such as the mismatch hypothesis (e.g., Ho et al., 2008; MacDorman et al., 2009), early suggestions and debate about possible underlying mechanisms of the uncanny effect (e.g., reviewed in MacDorman et al., 2009), compilation of available data and ideas (Pollick, 2009), introduction of empirically well-founded psychological methods and concepts to investigate the UVH (e.g., Cheetham et al, 2011; Saygin et al, 2011; Looser and Wheatley, 2011), and first signs of converging lines of thought and subsequent empirical and theoretical enquiry (e.g. Cheetham et al, 2013; Saygin et al, 2013; Looser and Wheatley, 2011, Yamada et al, 2012; Rhodes, 2012). Reflecting the early stages of enquiry and uncertainty as to concepts and measurement, this processes has include empirical studies favouring and studies failing to substantiate the notion of the uncanny effect (e.g., Green et al., 2008; MacDorman, 2005; MacDorman et al., 2009; Tinwell & Grimshaw, 2009; Hanson, 2006; MacDorman & Ishiguro, 2006; Brenton et al., 2005, Bartneck et al., 2007).

There are a number of reasons for this inconsistency in findings to date. These largely stem from one common problem: Mori did not embed his insightful ideas in any kind of scientific empirical framework within which to specify his terminology and to operationally define his conjectures for empirical testing. The UVH provides only a vague though illustrative description of a number of loosely related behaviours, feelings and affective evaluations. Mori refers to these collectively as *shinwakan*. This is an ambiguous Japanese neologism, the actual meaning of which is not clear. There have therefore been various renderings of it in translation, with *shinwakan* being investigated in terms of affective dimensions such as likability, pleasantness, and familiarity (i.e., familiar vs. strange) (e.g. Green et al., 2008; Seyama & Nagayama, 2009, 2007; Dill et al., 2012; MacDorman & Ishiguro, 2006; Tinwell &

Grimshaw 2009; Bartneck et al., 2007; Tinwell et al., 2011). Reflecting this *construct problem*, Mori (2012) recently and vaguely re-defined shinwakan as affinity. Nevertheless, the UVH speculates that shinwakan has a negative (i.e., the uncanny effect, or *bikumi* as Mori calls it) or positive character as a function of physical humanlike realism in appearance and motion. This speculation does convey a sense of the general problem that Mori envisaged. But its formulation in pre-scientific (everyday) and ambiguous terms that lack clear construct and operational definitions of shinwakan and bikumi as well as the lack of a clear specification of the relationship between these and humanlike realism renders the UVH unamenable to empirical testing without augmenting the UVH with further assumptions, or auxiliary hypotheses (Poundstone, 1988).

These auxiliary assumptions are needed in order to deal with a second likely reason for inconsistencies in findings to date. The appeal of the UVH is that it describes the notion of positive and negative subjective responses to humanlike objects in a simple and straightforward way. But the description does not reflect the likely complexity of the phenomenon the UVH seeks to characterize (Hodgins et al., 2010; Tinwell et al., 2009). From the perspective of psychology and cognitive neuroscience, the positively or negatively valenced response to and experience of humanlike entities (e.g., empathic feelings, emotional engagement, approach and avoidance behaviors, experience of eeriness, uncanny feelings) will be the result of a complex interplay of perceptual, cognitive and affective mechanisms and underlying neural processes (see e.g., Cheetham et al., 2009). This complexity is hinted at in initial neuroimaging studies of the relationship between perceived human likeness and neural processing in various cognitive domains and in the experimental paradigms used in these studies to disentangle and isolate components of the perception and experience of humanlike entities (Looser et al., 2012; Cheetham et al., 2011; Saygin et al., 2011; Chen et al., 2010; Chaminade et al. 2010; Krach et al., 2008; Hegel et al., 2008; Gazzola et al., 2007; Chaminade et al., 2007; Tai et al., 2004).

While multiple psychological mechanisms will underpin this experience and require, therefore, the application of various methodological strategies and paradigms to assess this experience, uncanny-related research has almost exclusively relied on the use of *ad hoc* developed self-rating scales to assess subjective experience (e.g.,

Tinwell et al., 2010; MacDorman et al., 2009a, 2009b; Seyama and Nagayama, 2009; Green et al., 2008; Ho et al., 2008; MacDorman, 2006; Hanson, 2006; see also Looser and Wheatley, 2011). The methodological reliance on self-rating scales has been favoured because these are inexpensive and easy to administer (see e.g., Ho et al., 2007). But the lack of construct specification in the UVH has left researchers faced with the general complexity of affective experience itself and its measurement (Nielsen and Kaszniak, 2007). A first real step toward alternative assessment strategies has been undertaken by Ho and MacDorman (2010), who constructed a multi-axes set of indices to capture subjective perception of humanness and experience of the uncanny effect, though the process of psychometric validation does not appear to be concluded (cf. Lienert and Raatz, 1994). Critically, the heavy reliance on self-rating scales for evaluation of subjective state is that the cognitive-evaluation process needed to for this is likely to interrupt the very pre-reflective experience of the humanlike entity that is of interest to the designer (Fahrenberg et al., 2007; Hegel et al., 2009; Eyssel et al., 2010).

The reliance on one method neglects the potential advantage of using convergent (and non-convergent) findings that might result from the use of other methods as a means to informing the iterative process needed to refine and specify the UVH' conceptualisation and measurement. Given that the UVH is not an empirically testable statement of a problem (cf. Hempel 1945) and perhaps better described as a non-scientific statement, pseudo-hypothesis or as conjecture (Popper, 2004), this iterative process would feed the development of scientific theory. The purpose of scientific theory would be to furnish the UVH with a coherent framework in which observed or experienced phenomena are clearly described on the basis of empirical evidence (Gioia and Pitre, 1990). This might extend beyond any evidence that substantiate the occurrence of uncanny feelings as a function of anthropomorphic realism (i.e., the uncanny effect) toward the generation of a well-developed uncanny valley theory that would incorporate predictive and explanatory knowledge as to when and why these feelings occur (e.g., Van de Ven, 1989; Gioia and Pitre 1990; Lynham, 2000, 2002). At present, the development of scientific theory in relation to the UVH has focussed on proposing theoretical accounts of the uncanny effect.

These explanatory accounts (reviewed in MacDorman et al., 2009) might be understood as implying that the uncanny effect is empirically well established (as a pre-condition for their explanation). This is in fact presumed in uncanny-related research suggesting how the designer should proceed to ensure that the humanlike character does not “fall” into the uncanny valley (e.g., MacDorman et al., 2009a, 2009b; Walters et al., 2008: for tool development to aid such design decision making, see Ho & MacDorman, 2010). Similarly, the uncanny effect has been used in empirical studies to explain the presence of unexpected findings and of participants’ negative responses to the appearance and behaviour of avatars or robots (e.g., Yamamoto et al., 2009; Aylett, 2004; Wages et al., 2004).

But these explanatory accounts might also be viewed as exploratory attempts to explicate the meaning of the ambiguous UVH. *Threat avoidance* is one such account that is already hinted at in the UVH. One variant of this account that largely bypasses without reference to it the wealth of psychological literature on threat avoidance (e.g. Gray, 1982) is the *avoidance of pathogens*. This suggests that cues that may be interpreted as indicative of disease are thought to elicit feelings of unease. The exploratory nature of such accounts can be seen in subsequent empirical studies. For example, Hodgins et al. (2010) and Tinwell et al. (2011) adopted and examined this particular account by altering in their studies the animation of avatar’s facial features to generate the impression of muscular disease. Similarly, the *mortality salience* account (MacDorman, 2005), based on terror management theory, hypothesizes that feelings of unease are associated with the confrontation with death and with the awareness of our own mortality. This account has been used for example in a study in which videos of horror game characters (e.g., vampires) were compare against photo-realistic and cartoon avatars (Tinwell and Grimshaw, 2009).

A further approach to the examination of the proposed uncanny effect that has received little attention is based on cognitive evaluations and attribution of mind, especially, concerning the perception of robots expressing a capacity for affective experience. This work revolves around the social cognitive *Theory of Mind* approach to ‘understanding others’. This entails the adoption by the observer of the other and explicit inference or evaluation of the mental state and feelings of the other (e.g. Fan et al. 2011). The initial findings of this approach indicate that robots are evaluated in

terms of the dimensions of agency and experience (Gray et al., 2007). Following this, Gray and Wegner (2012) lend support to the general notion of the UVH by showing that feelings of unease are induced when a machine is perceived to have the capacity to experience (i.e., to feel and to sense) but not agency (for the impact of inferential processing on experience of nonhuman entities, see also e.g., Waytz and Norton, 2014; Epley et al., 2008). Theory of Mind related processes have been investigated in connection with the UVH in neuroimaging research that reveal the recruitment of brain areas that subserve these evaluative mechanisms (e.g., Chaminade et al., 2007; Saygin et al., 2011). Processes of empathy that entail automatically sharing and implicitly generating a representation of the other entity's affective state (e.g., Decety and Jackson, 2004) have as yet received almost no attention, possibly with the exception of (MacDorman et al., 2013). Motivated by Mori's (2012) re-definition of shinwakan as empathy and affinity, MacDorman and colleagues' undertook to examine the UVH in terms of empathy-related processing. They were unable to substantiate this new perspective.

Uncanny-related research has largely focussed on the affective experience of humanlike entities (e.g., MacDorman et al., 2013; Green et al., 2008; Seyama & Nagayama, 2009, 2007; Dill et al., 2012; MacDorman & Ishiguro, 2006; Tinwell & Grimshaw 2009; Bartneck et al., 2007; Tinwell et al., 2011; Minato et al., 2006; Fabri et al., 2004; Walters et al., 2008; MacDorman et al., 2009a; Ho and MacDorman, 2010; Hanson, 2006; MacDorman, 2006; Tinwell and Grimshaw, 2009). But the way in which human likeness along the DHL is subjectively perceived and cognitively represented has received no attention. This question is critical for UVH because the UVH and uncanny-related researchers seeks to understand and to define subjective experience (i.e., shinwakan) as a function of perceived human likeness. The conceptual and experimental representation of the DHL is therefore of particular importance (Cheetham and Jancke, 2013).

Consideration of the UVH' conceptualisation of the DHL raises a number of issues. One way of understand the UVH and the illustrated curve is to view the shape of the curve and the objects depicted along it as a *visual hyperbole* (Roberts and Kreuz, 1994). The purpose of the exaggerated valley in the shape of the curve and the objects named or depicted along it is to render meaning to the basic ideas presented

in the UVH. Arguably, the illustration is not intended to invite literal interpretation. But the influence of the UVH' conceptualisation of the DHL is such that some studies have been guided by the illustrative depiction of the UVH (e.g., Tinwell and Grimshaw, 2009). These have used wide a range of stimuli that can be defined as varying along a perceptual dimension like human likeness. But there has been no control for visual cues of other perceptual dimensions that could confound with the processes under investigation, such as judgments of human likeness, perceptual discrimination performance, mind attributions, category judgements, or familiarity judgements along the DHL. For example, cues that relate to ethnicity, gender, facial distinctiveness, familiarity and identity, and facial expression confound with human likeness (Ho et al, 2008; Seyama and Nagayama, 2007; Hanson, 2006; MacDorman and Ishiguro, 2006).

Another issue is that the UVH conceptualizes the DHL as a linear dimension of physical similarity space, the human end of which is represented by a single human category exemplar. Clearly, Mori focuses on the wide variation of potential physical forms of nonhuman objects. But the idea that the human image might be treated as a general point of reference, as adopted in investigations and theoretical consideration of the UVH (e.g., Tinwell and Grimshaw, 2009; Ramey, 2005; Walters et al., 2008; Minato et al., 2006), implicitly assumes that variation in human likeness cannot be plotted along the DHL, that is, that human likeness does not vary within the human category. For this reason, the second positive peak in the hypothesis' original familiarity-human likeness relationship is located at the human end of the DHL (see Fig.1). The emphasis on the nonhuman aspect of the DHL has been influential in studies guided by the hypothesis (e.g. Tinwell et al., 2011; Tinwell, 2009; Schneider et al, 2007). But this assumption is incorrect (e.g., Cheetham et al., 2011; Looser and Wheatley, 2011; Yamada et al., 2012).

The preceding issue highlights one drawback of UVH-related research based on robots. The robot-based approach effectively places the focus of investigation of uncanny experience and subjective evaluations on entities that are clearly nonhuman. This in part because the researchers are specifically interested in the application of robots in different real-world contexts that are by design stylizations of the human 'model' (e.g., Walters et al., 2008; Minato et al 2008; Ho et al., 2008;

Green et al., 2008; Nowak & Biocca, 2003; Hegel et al., 2010). While some current robots do have features that permit a diversity of possible modulations in appearance, social expressiveness and biomimetic movement (e.g., Hegel et al., 2010; Hegel et al., 2006; RoboThespian, <https://www.engineeredarts.co.uk>), technological constraints on robotic design limit their use as a means to investigating the UVH. One essential reason for this is the point that the UVH is concerned – in its vague formulation - with the degree of nonhuman and human object similarity along the DHL at which difficulty making the distinction between nonhuman and human is most likely to evoke personal uncertainty and discomfort (i.e. at the point along the DHL characterised by the valley, see Fig 1). Empirical examination of the UVH requires therefore use of technology that allows experimentally controlled fine-grained manipulation of humanlike appearance and behaviour (Cheetham and Jancke, 2013). The fine-grained approach is important because the UVH does not predict the actual degree of perceived human-like appearance or behaviour at which the transition between positively and negatively valenced emotions and cognitions should occur.

One way to represent human likeness along the DHL is to generate morph continua (e.g., MacDorman and Ishiguro, 2006; Hanson et al., 2005; Ho et al, 2008; Seyama and Nagayama, 2007; Hanson, 2006). The specific advantage of computer-graphics technology is that it permits experimentally controlled fine-grained manipulation of humanlike appearance and behaviour to be brought into relationship with behavioural measures of subjective perception and experience. While static and dynamic stimuli would aid investigation of the contribution of invariant (e.g., facial texture, facial configuration) and variant (e.g., emotional expression) information to perceptual, affective and cognitive processing of human like entries, as reflected in various studies of affect and facial and whole body processing (e.g., Karnadewi and Lipp, 2011; Cloutier et al., 2008; Schindler et al., 2008), uncanny-research using computer-graphics has focussed on the use of static stimuli. This reflects the interest in Mori's suggestion that dynamic and non-dynamic stimuli or stimulus features modulate subjective experience differently and therefore the current focus on non-dynamic perceptual features.

The use of morph continua to represent the DHL has been subject to confounding factors that have limited the internal validity of studies aiming to test the UVH. These confounds include the use of two or even three different juxtaposed continua to represent the DHL (MacDorman and Ishiguro, 2006; Hanson, 2006). Though conducted to explore the uncanny effect described in the UVH, these juxtaposed continua effectively alter the UVH concept of human likeness and the ability to actually test the UVH by introducing perceptual discontinuities to the DHL (cf., Pastore, 1987). This means for example that equivalent increments of physical change along the DHL are not ensured. A central feature of the UVH is that perceptual difficulty discriminating between highly realistic humanlike entities and their human equivalent will evoke an unpleasant affective state. Controlling for physical differences between these entities along continua is therefore important for testing the UVH. A lack of experimental control is otherwise apparent in morphing artefacts resulting from the morphing of facial information like upper facial features and hair profile (Ho et al., 2008), using dissimilar continua parent images in which non-facial information such as head attire and facial jewellery are only present in one stimulus (Seyama and Nagayama, 2007), in the simultaneous presentation of human and nonhuman perceptual information in the same stimuli intended to represent the human end of the DHL (Hanson et al., 2005), and in morphing noise (i.e., disparities in the alignment of facial features between successive morphs of a continuum) in recent research (Yamada et al., 2012).

Mori did not pursue his general notion of the uncanny effect any further or subject his hypothesis to empirical examination. The UVH and the problem that it poses can be understood in its pre-scientific formulation as representing the initial stage of scientific enquiry (Northrop, 1966). At this stage, it can be assumed that Mori had not developed the scientific hypotheses needed to examine the problem that he conceived of. Exploratory thought is thus needed to specify more clearly the details and scope conditions (Cohen, 1989) of the UVH in such a way that the problem or aspects of it may be developed as testable hypotheses (Hempel, 1945). To this end, Mori's visual hyperbole might be better conceived of as a useful heuristic device rather than the object of research itself. As such, this device renders the ideas in the UVH accessible to exploratory thought. This thought requires the use of background knowledge from psychology and other domains of science in order to introduce to the

UVH established terminology, concepts, and methods as a basis for hypothesis testing and for interpretation of findings (Fleck, 1979; Hempel, 1945; Northrop, 1966; Poundstone, 1988). For example, the well-established concepts and knowledge on perceptual and category processing, perceptual decision making and *category perception* (CP) (e.g., Harnad, 1987) can be used to develop and specify testable ideas. The aim of the three studies described in the following sections (and the total of 7 experiments therein) was to introduce such concepts and knowledge to the investigation primarily of the DHL.

3. Study A. The human likeness dimension of the "uncanny valley hypothesis": behavioural and functional MRI findings

3.1. Abstract

The uncanny valley hypothesis (Mori, 1970) predicts differential experience of negative and positive affect as a function of human likeness. Affective experience of humanlike robots and computer-generated characters (avatars) dominates “uncanny” research, but findings are inconsistent. Importantly, it is unknown how objects are actually perceived along the hypothesis’ *dimension of human likeness (DOH)*, defined in terms of human physical similarity. To examine whether the DOH can also be defined in terms of effects of categorical perception (CP), stimuli from morph continua with controlled differences in physical human likeness between avatar and human faces as endpoints were presented. Two behavioural studies found a sharp category boundary along the DOH and enhanced visual discrimination (i.e., CP) of fine-grained differences between pairs of faces at the category boundary. Discrimination was better for face pairs presenting category change in the human-to-avatar than avatar-to-human direction along the DOH. To investigate brain representation of physical change and category change along the DOH, an event-related fMRI study used the same stimuli in a pair repetition-priming paradigm. Bilateral mid-fusiform areas and a different right mid-fusiform area were sensitive to physical change within the human and avatar categories, respectively, whereas entirely different regions were sensitive to the human-to-avatar (caudate head, putamen, thalamus, red nucleus) and avatar-to-human (hippocampus, amygdala,

mid-insula) direction of category change. These findings show that Mori's DOH definition does not reflect subjective perception of human likeness and suggest that future “uncanny” studies consider CP and the DOH’s category structure in guiding experience of nonhuman objects.

3.2. Introduction

The uncanny valley hypothesis (Mori, 1970) proposes that observation of a humanlike object (e.g., industrial robot, lifelike prosthetic hand, human corpse, doll, mannequin) can evoke positive or negative feelings and cognitions (referred to here as valence) depending on the object’s degree of physical similarity to human appearance (for recent overviews, see e.g., MacDorman et al., 2009; Pollick, 2006). Importantly, the relationship between valence and human likeness is suggested to be non-linear. As illustrated in Fig. 1, valence increases positively with greater human likeness up to a level of relatively high realism at the first peak of the curve along the dimension of human likeness (DOH). Positive valence reflects the experience of emotional engagement with and feelings of empathy for the humanlike object. At greater degrees of realism, the observer encounters difficulty in distinguishing an object from its natural human counterpart and experiences personal discomfort. Mori characterises this discomfort as an uncanny feeling marked by a sense of strangeness, eeriness and disquiet that can extend to feelings of disgust and revulsion. This uncanny feeling is reflected in a sharp negative peak or valley (i.e., uncanny valley) in the slope of the depicted valence-human likeness relationship. When an object’s appearance is so realistic that it is perceived to be human the valence of associated affect and cognition is thought to reach a second positive peak.

The hypothesis’ central prediction that individuals can feel less emotionally engaged or even distracted by relatively realistic humanlike objects has been very influential in guiding animators, video game designers, and roboticists in the design of virtual characters (e.g., Fabri et al., 2004) and robots (e.g., Minato et al., 2006). This has lead to research into how characters should be designed to avoid “falling” into the uncanny valley (e.g., MacDorman et al., 2009a, 2009b; Walters et al., 2008) and into tool development to aid such design decision making (Ho & MacDorman, 2010). The

idea of the uncanny valley has also been used in empirical studies to account for unexpected findings and negative responses of participants to the appearance and behaviour of avatars or robots (e.g., Yamamoto et al., 2009; Aylett, 2004; Wages et al., 2004). In view of this interest, it is noteworthy that Mori did not subject his working hypothesis to empirical examination. In fact, research of the valence-human likeness relationship is in its infancy, largely focussing on the valence of subjective feelings and evaluations in response to variously realistic nonhuman characters. But findings have been inconsistent (e.g., Tinwell et al., 2010; Tinwell & Grimshaw, 2009; MacDorman, 2006; Hanson, 2006), this being in part attributable to the uncertainty surrounding Mori's vague definition of the valence dimension (e.g., MacDorman et al., 2009a; Seyama & Nagayama, 2007). In contrast, the way in which human likeness along the DOH is actually perceived has not been scrutinised.

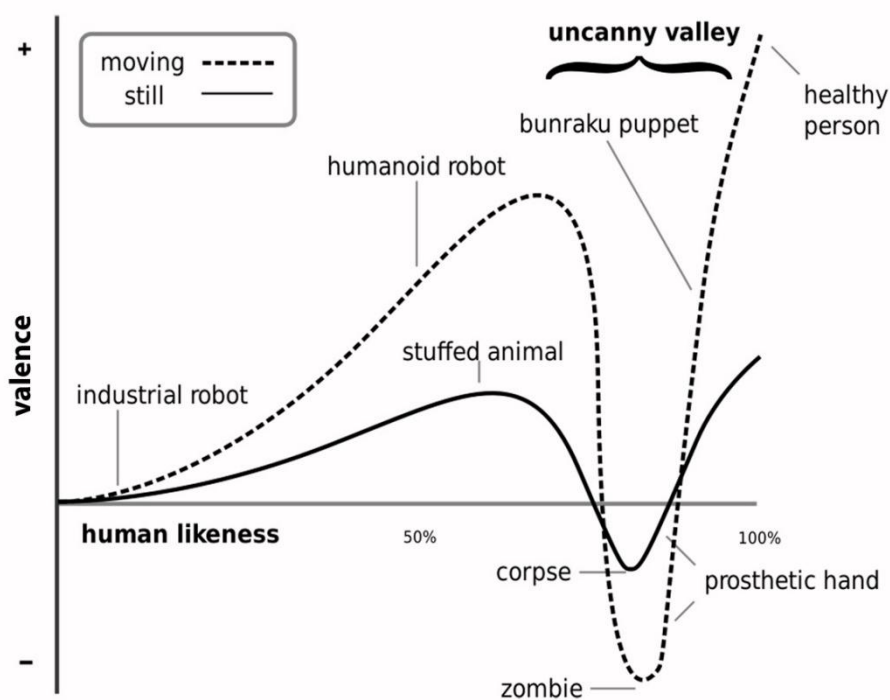


Figure 1. Illustration of the non-linear relationship between the experience of negative and positive affect (valence) and perceived human likeness. The otherwise positive relationship shows a sharp negative peak (i.e., uncanny valley) at the level of realism at which subtle differences in the appearance and behavior of a highly realistic yet discernibly unnatural humanlike object is suggested to elicit a sense of strangeness and personal discomfort (i.e., an uncanny feeling). Illustration adapted from MacDorman (2005b).

The hypothesis' central prediction that individuals can feel less emotionally engaged or even distracted by relatively realistic humanlike objects has been very influential in guiding animators, video game designers, and roboticists in the design of virtual characters (e.g., Fabri et al., 2004) and robots (e.g., Minato et al., 2006). This has lead to research into how characters should be designed to avoid "falling" into the uncanny valley (e.g., MacDorman et al., 2009a, 2009b; Walters et al., 2008) and into tool development to aid such design decision making (Ho & MacDorman, 2010). The idea of the uncanny valley has also been used in empirical studies to account for unexpected findings and negative responses of participants to the appearance and behaviour of avatars or robots (e.g., Yamamoto et al., 2009; Aylett, 2004; Wages et al., 2004). In view of this interest, it is noteworthy that Mori did not subject his working hypothesis to empirical examination. In fact, research of the valence-human likeness relationship is in its infancy, largely focussing on the valence of subjective feelings and evaluations in response to variously realistic nonhuman characters. But findings have been inconsistent (e.g., Tinwell et al., 2010; Tinwell & Grimshaw, 2009; MacDorman, 2006; Hanson, 2006), this being in part attributable to the uncertainty surrounding Mori's vague definition of the valence dimension (e.g., MacDorman et al., 2009a; Seyama & Nagayama, 2007). In contrast, the way in which human likeness along the DOH is actually perceived has not been scrutinised.

While Mori defined the DOH in terms of a linear change in the degree of physical humanlike similarity, we proposed that the subjective perception of objects along the DOH might be described differently, namely, in terms of the effects of categorical perception (for CP see, e.g., Harnad, 1987). Applied to the DOH, CP means that irrespective of physical differences in humanlike appearance objects along the DOH are treated as conceptually equivalent members of either the category 'nonhuman' or the category 'human', except at those levels of physical realism at the boundary between these two categories. At this category boundary, the available sensory evidence does not allow rapid and effortless discrimination of an object in terms of the observer's category representations of nonhuman and human. Consistent with this, ambiguity in discriminating a humanlike object from its natural human counterpart lies at the heart of Mori's hypothesis, and, when no such ambiguity is experienced, Mori implicitly assumes that the observer assigns his or her sensory impressions of an object to the nonhuman or the human category. The defining

feature of CP is considered to be enhanced discrimination of pairs of different stimuli that are perceptually adjacent (along a dimension such as the DOH) but straddle opposite sides of the category boundary (such as between the categories human and nonhuman) and poorer discrimination of pairs of different stimuli from within a given category (Pastore, 1987). In other words, the ability to discriminate between physically different stimuli might not be the same at all points along the DOH, with enhanced sensitivity for objects closest to the category boundary (for CP criteria, see e.g., Studdert-Kennedy et al., 1970).

It is important to note that CP does not occur along any human likeness dimension. Campbell et al. (1987) found CP for faces of humans and cows but not for humans and monkeys. CP and indeed processing of the human category along Mori's DOH has not been investigated. In fact, a prominent feature of Mori's hypothesis is that it does not consider the possibility that objects assigned to the human category might also differ in the degree of humanlike similarity along the DOH. Our interest in understanding the categorical structure of the DOH is based on the assumption that negatively valenced or uncanny experience is likely to occur in association with categorisation ambiguity for stimuli at or closest to either side of the nonhuman-human boundary. For this, the human category and the potential impact of CP on processing objects along the full length of the DOH need to be considered. The present investigation focused therefore on the subjective perception of human likeness rather than on valence in order to provide a clear theoretical-psychological framework for further studies of uncanny experience.

Two behavioural and one neuroimaging study was conducted. The behavioural studies aimed to demonstrate that subjectively assigned human faces do vary in humanlike appearance within the human category along the DOH and to test the proposal that subjective perception of human likeness along the DOH shows effects of CP. To represent the DOH, linear continua of morphed faces were generated using avatar and human faces as endpoints to create a controlled transition of physical similarity between them. Using stimuli drawn from these continua, the first study entailed a forced-choice classification task to determine the presence and location of the avatar-human category boundary, and the second used a perceptual discrimination task to determine the presence and location of the discrimination

boundary (i.e. the point of enhanced perceptual discrimination) and thus to verify CP. The subsequent neuroimaging study used functional magnetic resonance imaging (fMRI) to investigate nonhuman-human category processing in the brain. The first aim of this was to show that distinctly different brain regions are responsive to processing the physical similarity of faces along the DOH and to processing the DOH's discrete human and nonhuman categories. For this, we used a pair-repetition priming paradigm (for reviews see, Henson, 2003; Grill-Spector & Malach, 2001). This paradigm entailed the presentation of stimulus trials, each comprising a pair of faces (i.e. prime and target) displayed in quick succession, and the measurement of the physiological BOLD response to the primed targets as an indicator of neural activity. Repetition in the target of information presented in the prime has been shown to induce neural adaptation (i.e. attenuation of neural activity) in brain areas selectively responsive to processing the repeated information (e.g., Grill-Spector et al., 2006; Jiang et al., 2006). This neural adaptation effect and its regional localisation have been successfully used to disentangle regions responsive to physical and category change (e.g., Jiang et al., 2007; Rotshtein et al., 2005).

We focussed in the fMRI study on category change, hypothesising that implicit processing of the discrete change in category between our prime and target stimuli would modulate neural activity in regions associated with category learning and categorisation uncertainty, that is, basal ganglia, medial temporal lobe, thalamus, medial frontal gyrus, and the anterior insula (e.g., Fleming et al., 2010; Seger & Miller, 2010; Li et al., 2009; Ashby & Maddox, 2005; Grinband et al., 2006). Mori focussed in his informal description of the hypothesis on the situation in which a nonhuman object is initially mistaken for human. The second aim of this fMRI study was to explore the possibility that the anticipated effects of category change in these brain regions might be different depending on the actual direction of category change between the stimuli used as prime and target, that is, in the human-to-avatar direction along the DOH (i.e., human as prime and avatar as target) or the avatar-to-human direction (i.e. avatar as prime and human as target). The basis for this idea was that different brain regions (e.g. in the basal ganglia and medial temporal lobe) are known to be differentially modulated depending on categorisation experience with a given category (e.g., Poldrack et al., 1999). Given the asymmetrical category knowledge of our participants with human and novel nonhuman faces, we assumed that this

differential effect might similarly apply for category processing along the DOH in individuals with little or no experience of our nonhuman stimuli. Category-related processing of highly realistic and potentially biologically salient human-like faces might also modulate neural responses in further regions specifically associated with appraisal of affective valence (e.g., Vuilleumier, 2005) or processing under conditions of valence ambiguity (e.g., Herwig et al., 2007). We explored this, anticipating involvement of regions commonly associated with affective processing, especially the amygdala (e.g., Todorov & Engell, 2008; Herwig et al., 2007).

3.3. Material and methods

Participants

Healthy male and female, consistently right-handed (Annett, 1970) adult volunteers with no record of neurological or psychiatric illness and no current medication use volunteered for one of the three studies. All participants were students of the University of Zurich, native speakers of Swiss-German or Standard German, reported having no explicit experience with avatars such as those used in video games, virtual role playing games, or second life, and had normal or corrected-to-normal vision. Written informed consent was obtained before participation according to the guidelines of the Declaration of Helsinki. Each volunteer received 20 Swiss Francs for participation. The study and all procedures and consent forms were approved by the Ethics Committee of the University of Zurich.

Stimuli

Avatar-human morph continua were generated to represent Mori's DOH. The selected stimuli for use as endpoints in the morphing procedure were 32 photographic images of unknown people and 32 facial images of avatars, together forming 32 morph continua (see Fig. 2.B). All faces were male, presented with full face, frontal view and neutral expression. The avatars were generated with the modelling suite Poser 7 (Smith Micro Software, <http://www.smithmicro.com>), which permits considerable texture detail and control over the facial mesh. The facial geometry and texture of the avatars were modelled (age and configural cues) to closely match the human counterpart to minimise perception of biological motion during quick successive presentation of faces (e.g. Schultz & Pilz, 2009). The images

were edited in Adobe Photoshop CS3, masking external features with an elliptic form and black background (72 dpi and 560 x 650 pixels). Contrast levels, overall brightness and skin tone of each pair of colour images of faces used as the endpoints of each continuum were adjusted to match before morphing. Morpher 3.3 (Zealsoft Inc., Eden Prairie, MN) was used to generate the linear morph continua. Each continuum comprised 13 different morphed images. These images or morphs were labelled M0 (beginning with the avatar endpoint) through to M12 (at the human endpoint), each morph representing a difference in physical appearance at increments of 8.33%.

All morphs (i.e., M0 to M12) were presented in the forced choice classification task of study 1. As described in the following, only those morphs representing increments of 33.33% in physical difference along each continuum (i.e., M0, M4, M8, M12) were used for the subsequent perceptual discrimination task of study 2 and the target-monitoring task in the (fMRI) study 3.

3.4. Experiment 1: Forced choice classification

A two-alternative forced-choice classification task was conducted to determine the presence of a sigmoid shaped response function along the avatar-human morph continua. This response function is considered to indicate a category boundary (Harnad, 1987).

Participants

N=25 volunteers (13 female, mean age 21.8 years; range 19 - 28 years) participated.

Materials and Procedure

All participants were tested individually. Each participant received written instructions presented on the screen before commencement of the experiment. A practice pre-test of 5 trials using stimuli from continua not included in the main test was then performed to ensure that the instructions had been understood. It was again ensured that participants understood the meaning of the word avatar. All morph stimuli were presented individually in random order for a total presentation of 416 trials. Each trial

began with the presentation of a fixation point for 500ms (participants were required to maintain fixation), followed by a morph image for 750ms. The participant was asked to identify the stimulus as either an avatar or human as quickly and precisely as possible by pressing one of two response keys. A black screen with fixation point remained after morph image presentation until the participant pressed a key, after which a blank black screen without fixation cross remained for 1500 ms until the next trial began. The task was conducted in a sound attenuated and light-dimmed room, and morph stimuli were presented on a LCD monitor (1280 x 1024 resolution, 60 Hz refresh rate, at eye-to-monitor viewing distance of 62 cm), using Presentation® software (Version 14.1, www.neurobs.com)

Analyses

The classification data were summarised by the shape of the avatar-human classification curve as described by the slope of the response function. The slope was determined by fitting logistic function models to the response data of each participant and continuum, and the parameter estimates were derived. Individual continua were analysed across participants to ensure best fit of logistic functions, these reflecting a step-like shape in the avatar-human classification curve and thus a categorical component (Harnad, 1987). To test for a step-like shape across all continua, the derived parameter estimates for the logistic function of each continuum, averaged across participants, were tested against zero in a one-sample t-test. This and the average boundary value (i.e. the morph position associated with greatest decision uncertainty) of the fitted logistic curves are reported. SPSS 16 was used for data analysis.

3.5. Experiment 2: Perceptual discrimination

A variant of the same-different discrimination task (e.g., Angeli et al., 2008) was performed to determine CP along the morph continua. Participants judged whether the presented faces of a face pair were both the same or different in appearance. Greater performance accuracy in "different" judgements for face pairs that are perceptually adjacent (and therefore physically different along the avatar-human

continua) but straddle opposite sides of a category boundary (as determined in the preceding task) is taken as evidence of CP (Pastore, 1987).

Participants

N = 20 participants (9 female, mean age 25.1 years; range 18 - 30 years) were tested.

Materials and Procedure

The choice of continua for use in this task was determined on the basis of the results (see Results section 3.1.) of the preceding forced choice classification task. Four morphs M0 and M4 (categorised as avatars) and M8 and M12 (categorised as human) were selected from each of the continua (see Fig. 2B), with M4 and M8 straddling the category boundary. M0, M4, M8, and M12 represent increments of 33% along the length of each continuum to ensuring control of physical dissimilarity. For the perceptual discrimination task, stimulus trials were presented that each comprised a pair of faces (see Fig. 3). The first face of a face pair was either an avatar or a human. Trials in which the first face was an avatar are referred to as 'avatar' trials (see Fig. 3, C) and those in which a human face was presented first are referred to as 'human' trials (see Fig. 3, D). In the avatar trials, the first face was always the avatar morph M4, and the first face in the human trials was always the human morph M8. The choice of the second face of each face pair was determined according to three different face pair conditions: 'within' category face pairs (both faces drawn from within a category), 'between' category face pairs (first and second face morphs lying either side of the category boundary), and the 'same' stimulus face pairs (faces of a pair are identical). In the within category condition, the second face was morph M0 for avatar trials and M12 for human trials. In the between category condition, the second face was morph M8 for avatar trials and M4 for human trials. In the 'same' stimulus condition, M4 was used as the first and second face for avatar trials and M8 as the first and second face for human trials. Based on these conditions, the presentation of face pairs was therefore as follows: Avatar trials comprised the face pair morphs M4 - M0 for the within, M4 - M4 for the same, and M4 - M8 for the between conditions, and human trials comprised the face pair

morphs M8 – M12 for the within, M8 – M8 for the same, and M8 – M4 for the between conditions.

Both faces of each face pair were always drawn from the same continuum in which they were originally morphed. The presentation of face pairs was pseudo-randomised so that no trials of face pairs from within the same continuum were shown in close sequence. The presentation of avatar or human trials from a given continuum was determined randomly but counterbalanced across all participants in such a way as to ensure that each participant viewed either avatar or human trials from any given continuum, and that in total an equal number of avatar or human trials were viewed. Each face of a face pair was presented for 500ms with an inter stimulus interval (ISI) of either 75ms or 300ms between the faces of a pair. This was to determine whether ISI duration would differentially impact judgements (see, e.g., Rotshtein et al., 2005). Stimulus onset asynchrony (SOA) was 2,500ms between trials of face pairs. Please note that these features of the experimental design of stimulus presentation were also carried over to the fMRI investigation of Study 3 (and are therefore not described again in the Materials and Procedure section of Study 3). This experimental design allowed within category effects for faces from the avatar category and for faces from the human category to be examined, and the effects of category change between face pairs comprising an avatar and a human face to be examined in both directions along the DOH, that is, in the avatar-to-human and human-to-avatar directions.

The task was conducted in a sound attenuated and light-dimmed room, and morph stimuli were presented on a LCD monitor (1280 x 1024 resolution, 60 Hz refresh rate, at eye-to-monitor viewing distance of 62 cm), using Presentation® software (Version 14.1, www.neurobs.com)

Analyses

To examine discrimination accuracy for face pairs that crossed the category boundary compared with face pairs from the same side of the boundary, the 'different' responses (indicating the judgement that both faces of a pair were of different physical appearance) were computed as proportions of the total number of face trials and subjected to 3 x 2 ANOVA, with 3 face-pair conditions (within, between, same)

and 2 ISI (short and long). The data for avatar trials and human trials were treated separately in analysis. Greenhouse-Geisser adjustment was applied to correct the degrees of freedom whenever the assumption of sphericity was violated. SPSS 16 was used for data analysis.

3.6. Experiment 3: Target monitoring and fMRI

Participants

N = 22 volunteers (10 female; mean age, 21.9 years; range, 19-27) participated in the fMRI study.

Materials and Procedure

A pair-repetition paradigm with a target-monitoring task was applied. The stimulus conditions (i.e. the morph stimuli for the face pairs in the within, same and between conditions in the avatar and human trials, and the presentation times for the stimuli, ISI, and SOA) were the same as described in the preceding perceptual discrimination task.

The target-monitoring task required participants to press a response button upon detection of one of four possible up-turned rare target faces (M0, M4, M8 or M12) selected at random from an unused morph continuum. This was to ensure attention of the participants to the stimuli of interest. This task requires no explicit judgements of human or avatar category or physical similarity of faces, thus eliminating the confounding effects on BOLD signal modulation by motor response requirements (Henson 2003). Differential attention to prime and target was avoided in that the up-turned target face was presented as the first or second face of a face pair trial. There were 192 trials of face pairs, 15% of this trial number being in addition rare target faces and 25% in addition null events with fixation cross but no face stimuli. Participants were required to maintain fixation. Each scanning session consisted of two experimental runs. The visual stimuli were presented using the "VisuaStim - Digital" MRI-compatible head-mounted display (Resonance Technology Inc.), with a

visual mono display, resolution of 800 x 600, and 30° field of view. The total scanning time was approximately 26 minutes.

fMRI data acquisition

Structural and functional images were acquired using a 3T whole-body MR unit (Philips Medical Systems, Best, The Netherlands) and eight-channel Philips SENSE head coil. Structural images of the entire brain were registered using a T1-weighted three-dimensional, spoiled gradient echo pulse sequence (180 slices, TR = 20 ms, TE = 2.3 ms, flip angle = 20°, FOV = 220 mm x 220 mm x 135 mm, matrix size = 224 x 187, voxel size = 0.98 mm x 1.18 mm x 0.75 mm, resliced to 0.86 mm x 0.86 mm x 0.75 mm). Functional images were acquired from 225 whole-head scans per run using a Sensitivity Encoded (SENSE) (Pruessmann et al., 1999) single-shot echoplanar imaging technique (repetition time, TR = 2.6 s; echo time, TE = 35 ms; field of view = 220 mm x 220 mm x 132 mm; flip angle = 78°; matrix size = 80 x 80; voxel size = 2.75 mm x 2.75 mm x 4 mm, resliced to 1.72 mm x 1.72 mm x 4 mm). Susceptibility artefacts in temporal cortices were reduced by adjusting the slice tilt to 30° from the transverse plane (Weiskopf et al., 2007). Three dummy scans at the beginning of each run were acquired and discarded in order to establish a steady state in T1 relaxation for all functional scans.

fMRI data analysis

Preprocessing and MRI data analysis were performed using MATLAB 2006b (Mathworks Inc., Natick, Massachusetts, USA) and the SPM5 software package (<http://fil.ion.ucl.ac.uk/spm>). All images were realigned to the first recorded volume, normalised into standard stereotactical space (using the EPI-template provided by the Montreal Neurological Institute, MNI brain), resliced to 2 mm x 2 mm x 2 mm voxel size and smoothed using a 6-mm full-width-at-half-maximum Gaussian kernel. Activated voxels were identified by a general linear model (Friston et al., 1995) implemented in SPM5. High-pass filtering (cut-off 128 s) was applied to the time series. For each subject, the fMRI responses were modelled with a design matrix using the onset of the second face of each face-pair in each of the six face pair conditions (i.e., within, same and between for avatar and human trials) and the onset

of the target events (for target monitoring) as regressors convolved with SPM's standard canonical hemodynamic response function (HRF). The parameter estimates of the HRF for each regressor were calculated for each voxel, and linear contrasts were computed for each subject (Friston et al., 1995). These contrasts were entered into a second-level model. All contrast images were first smoothed using a 8 mm FWHM Gaussian kernel to allow inter-subject localisation differences to be accounted for, with a final estimated overall smoothing of 10 mm FWHM $((6^2+8^2)^{1/2})$. One-sample random effects t-statistics across subjects was performed to allow for population inferences.

The pair-repetition paradigm used the same face pairs and face-pair conditions as described for the preceding perceptual discrimination task. Briefly, the 'within', 'same' and 'between' face pair conditions used the face pair morphs M4 – M0 (i.e., within), M4 – M4 (i.e., same), and M4 – M8 (i.e., between) in the avatar trials and M8 – M12 (i.e., within), M8 – M8 (i.e., same) and M8 – M4 (i.e., between) in the human trials (see Fig. 3,B). Contrasts were defined on the basis of these face-pair conditions in order to identify brain regions sensitive to physical and category-related change between the presented prime and target stimuli. The rationale behind this is that (further to the description in the introduction) the neural adaptation effect (Grill-Spector & Malach, 2001) evokes a smaller BOLD signal in response to the second of a pair of identical stimuli than to the second of a pair of dissimilar stimuli. As the targets in the same, within and between conditions represent dissimilar points along the DOH in terms of physical attributes and/or category, relative differences in signal decrease between the face-pair conditions permits localisation of neuron populations responsive to change in physical attributes and/or category along the DOH (see e.g., Jiang et al., 2009; Grill-Spector et al., 2006; Murray & Wojciulik, 2004). As the prime was always M4 in avatar and M8 in human trials, the following contrasts refer to the targets of each face pair condition in the avatar (M0, M4, M8) and human trials (M12, M8, M4). Sensitivity to physical change was detected using the contrast of conditions M0 (i.e., within) plus M8 (i.e., between) > M4 (i.e., same) for avatar trials and M12 (i.e., within) plus M4 (i.e., between) > M8 (i.e., same) for human trials. To detect brain regions selectively responsive to category change across the boundary in the direction avatar to human and human to avatar, the contrasts M8 (i.e., between) >

M4 (i.e., same) plus M0 (i.e., within) for avatar trials and M4 (i.e., between) > M8 (i.e., same) plus M12 (i.e., within) for human trials were used.

Whole-brain voxel-wise analyses was performed. In examining the sensitivity of brain regions to variation in physical attributes, we expected and focussed on activations in bilateral regions of the mid-fusiform gyrus (e.g., Xu et al., 2009; Jiang et al., 2009), but for purposes of comparison with other studies (e.g., Rotshtein et al., 2005) all voxels are reported that survived significance thresholding at $p < .001$, uncorrected for multiple comparisons with a spatial extent threshold of $k = 5$ voxels.).

The hypothesised regions of interest (ROIs) sensitive to category change are described in the introduction. The exploration of further regions associated with affective processing and affective ambiguity focused on the amygdala. Because of this dual approach with predefined ROIs and exploration of additional regions, all clusters of voxels ($k = 20$ voxels) responsive to category change that survived a threshold of $p < 0.005$ (uncorrected) are reported and discussed. The more lenient threshold (but more stringent level of contiguous voxels) reflected the exploratory interest in reducing type-2 errors (e.g. Wager et al., 2003; Phelps et al., 2008) coupled with the view that hemodynamic responses in sub-cortical structures of the affect processing network are considered more difficult to detect (e.g. Herwig et al., 2007; Etkin et al., 2006; Phelps, 2004). But, those voxels surviving significance thresholding at $p < .001$, uncorrected for multiple comparisons with a spatial extent threshold of $k = 20$ voxels, are also reported in the results section.

3.7. Results

Forced choice classification

The logistic slope value of the fitted regression curve of each individual continuum was highly significant at $p > 0.001$. Given the large number of values for all continua, we report the test of this response function across all continua and against zero in a one-sample t-test. This showed that the response function has a highly significant logistic component ($t_{31} = 29.28$, $P > 0.001$) that reflects a sigmoid step-like function consistent with the presence of a category boundary (Harnad, 1987). The grand

mean (see Fig. 2 A) of the fitted logistic curves (and of the response data) across continua shows the sigmoid-shaped curve with lower and upper asymptotes of avatar or human categorisation responses (as percentages of “different” responses) nearing 100% for avatars and 100% for humans, respectively. Across continua, the mean category boundary value ($M = 6.30$, $SD = 0.85$) corresponds with morph M6 (i.e. the midpoint along the morph continua). This value is derived from the fitted logistic curve and the ordinate midpoint between the lower and upper asymptotes of the categorisation responses, indicating therefore that the maximum uncertainty of 50% in categorisation judgements was associated with the morph M6.

Morphs M0, M4, M8 and M12 were selected for presentation in the subsequent two studies. The mean percentage of avatar or human identifications across all continua (reported here in terms of the response "human" as shown in Fig. 2 A) for the morphs M0, M4, M8 and M12 was 2.25 ($SD = 2.5$), 10.25 ($SD = 4.8$), 89.27 ($SD = 8.4$), 98 ($SD = 2.9$), respectively (see, Fig. 2 B).

In the interest of gaining an overall picture of categorisation response times (RT) along the DOH, an RM-ANOVA with factors continuum (32 continua) X morph position (13 levels) and RT as dependent variable was conducted. The analysis revealed no effect for continuum, but there was a main effect for morph position, $F(2.75, 66.01) = 27.04$, $p < 0.001$. Consistent with the preceding result of maximum uncertainty in decision responses at morph M6, the mean RT across participants indicated longest RTs for M6. To characterise this more clearly, the mean RT values at M6 were compared with the mean RT values at all other morph positions. A one-way RM-ANOVA analysis with morph position (2 levels: M6 versus all other morphs) and RT in seconds as dependent variable collapsed across continua showed that RT for M6 ($M = 1.42$, $SD = 0.26$) differed highly significantly from RT for the other morph positions ($M = 0.99$, $SD = 0.46$), $F(1,24) = 62.04$, $p < 0.001$.

In the interest of gaining an overall picture of categorisation response times (RT) along the DOH, an RM-ANOVA with factors continuum (32 continua) X morph position (13 levels) and RT as dependent variable was conducted. The analysis revealed no effect for continuum, but there was a main effect for morph position, $F(2.75, 66.01) = 27.04$, $p < 0.001$. Consistent with the preceding result of maximum

uncertainty in decision responses at morph M6, the mean RT across participants indicated longest RTs for M6. To characterise this more clearly, the mean RT values at M6 were compared with the mean RT values at all other morph positions. A one-way RM-ANOVA analysis with morph position (2 levels: M6 versus all other morphs) and RT in seconds as dependent variable collapsed across continua showed that RT for M6 ($M = 1.42$, $SD = 0.26$) differed highly significantly from RT for the other morph positions ($M = 0.99$, $SD = 0.46$), $F(1,24) = 62.04$, $p < 0.001$.

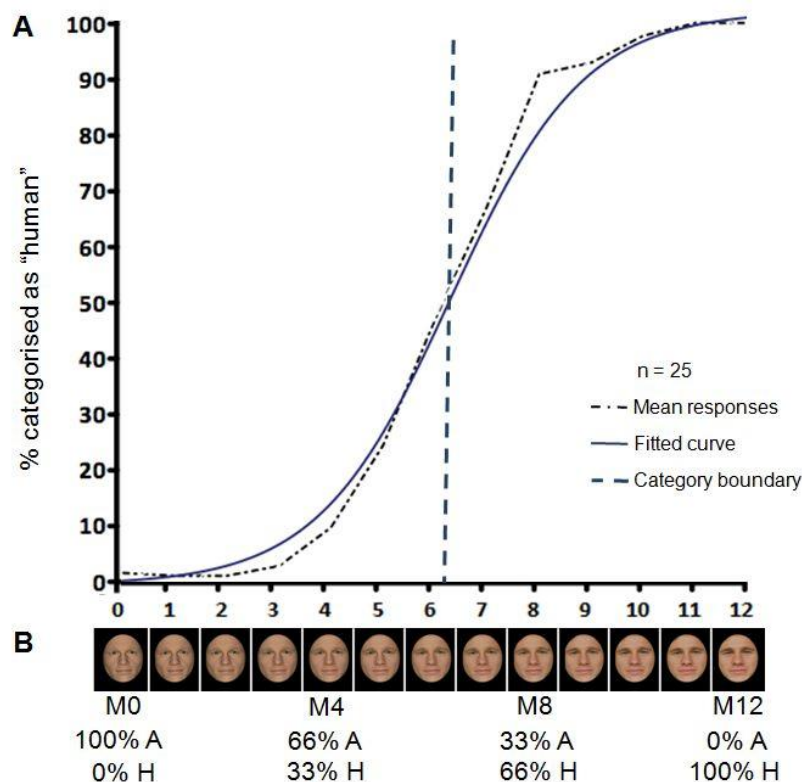


Figure 2. Results of the forced choice classification task (A) and example of a morph continuum (B). Mean percentage of responses (across 25 participants) in the forced choice classification task (A), showing the grand average of the response data (dashed black line), the fitted logistic response curve (solid blue line), and the estimated mean category boundary value (blue dashed line) over all continua. Results indicate a step-like response function consistent with the presence of a category boundary. Example of a morph continuum (B) with 13 levels from M0 to M12. The relative degree of linear physical transition between the avatar (e.g., 100 % A) and human (e.g., 0 % H) endpoints of morphs M0 and M4 (avatar category) and M8 and M12 (human category) are shown as percentages. The controlled physical dissimilarity of M0, M4, M8, and M12 was 33% along the continuum. These morphs were used in the subsequent perceptual discrimination task and fMRI study.

As the morphs M0, M4, M8 and M12 were to be used in the subsequent two studies, we tested also for differences in RT between these morphs. A one-way RM-ANOVA analysis with morph position (4 levels: M0, M4, M8 and M12) and RT as dependent variable collapsed across continua was conducted. The analysis revealed an effect for morph position, $F(1.77, 41.32) = 19.19$, $p < 0.001$. Tests of planned within-subject contrasts showed significant differences in RT within each category such that RT for M0 ($M = 0.67$ s, $SD = 0.27$) was shorter than for M4 ($M = 1.05$ s, $SD = 0.71$) [$F(1.24) = 48.2$ $p < 0.001$] and RT for M12 ($M = 0.91$, $SD = 0.56$) was shorter than RT for M8 ($M = 1.19$ s, $SD = 0.56$) [$F(1.24) = 17.2$ $p < 0.001$], indicating quicker responses for morphs at either end of the continua than for the morphs M4 and M8 that straddled the category boundary. But, there was no significant difference in RT between M4 and M8.

Perceptual discrimination task

Avatar trials

For the 'avatar' trials (see Fig. 3, A), the ANOVA analyses showed a significant main effect on discrimination accuracy of face pair condition, that is, within, same, and between [$F(2,38) = 149.74$, $p < .001$] and of ISI [$F(1,19) = 10.28$, $p = .006$]. The ISI effect appears to reflect a general bias irrespective of face pair condition toward more “different” responses in the short ISI condition than in the long ISI condition. Pre-planned tests of within-subject contrasts revealed that face-pairs that crossed the category boundary ('between' trial type) were significantly more often indicated as different than were those pairs from within a category ($F(1,19) = 142.89$ $p < .001$). There was also a significant effect of discrimination accuracy within the avatar category because pairs from within a category (i.e. 'within' trial type) were more frequently indicated to be different than those pairs of the "same" trial type, $F(1,19) = 6.09$, $p < .026$.

Human trials

Similarly, for the 'human' trials (see Fig. 3, B), the ANOVA analyses showed a significant effect on discrimination accuracy of face pair condition [$F(2,38) = 876.46$, $p < .001$] but no effect of ISI. Pre-planned tests of within-subject contrasts revealed

that face-pairs that crossed the category boundary ('between' trial type) were significantly more often indicated as different than were those pairs from within a category ('within' trial type), $F(1,19) = 932.03$, $p < .001$. There was also an effect of discrimination accuracy within the human category because pairs from within a category were significantly more frequently indicated to be different than those pairs of the 'same' trial type, $F(1,19) = 478.52$, $p = .028$.

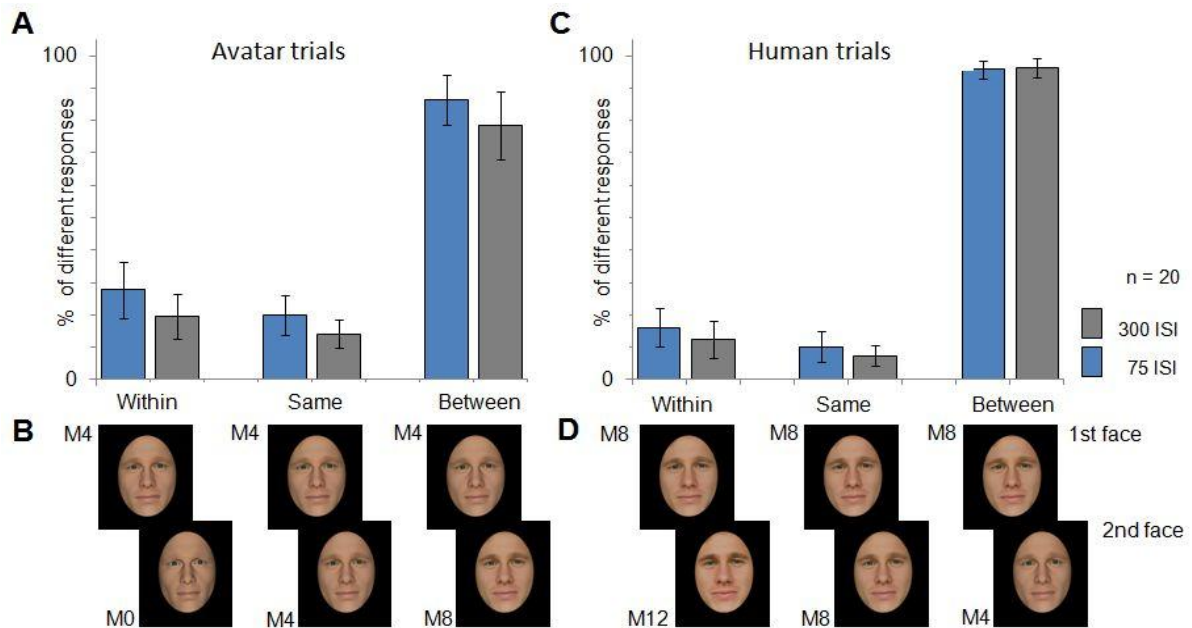


Figure 3. Results of the perceptual discrimination task (A,C) and example of stimulus conditions (B, D). The three different stimulus conditions for pairs of faces presented in the perceptual discrimination task and the fMRI study are shown (B, D): the faces of a pair were drawn from the within the same category (“within”), were identical (“same”), or showed category change by straddling opposite sides of the category boundary (“between”). In the perceptual discrimination task, same-different judgments were made according to whether faces were the same or different in physical appearance. Morphs M0, M4, and M8 were used for analysis of avatar trials (B) and M4, M8, and M12 for human trials (D): the first face in (B) is always M4 and in (D) always M8. The upper panels show the proportion of “different” responses for avatar (A) and human (C) trials across 20 participants. Controlling for relative distance of morphs along the continua, results show better discrimination accuracy for face pairs that crossed the category boundary (as determined in the forced choice classification task) than for pairs drawn from the same (avatar or human) side of the boundary, thus demonstrating categorical perception along our continua of human likeness for avatar and human faces.

In summary, the results demonstrate that processes of categorical perception are very likely to influence processing of objects along the DOH. The effect shown in this task was one of better discrimination accuracy for pairs that crossed the category boundary than for equidistant pairs drawn from within a category on one or other side of the boundary, this effect applying similarly for both avatar and human trials.

Target monitoring task and event-related fMRI

On this basis of the perceptual discrimination task, brain activity was expected to be differentially affected by the within and between category conditions. The ISI condition (short vs. long) of the perceptual discrimination task was carried over to the fMRI study.

Sensitivity to physical change

To detect brain regions sensitive to differences in physical features, face pairs in which there was a change in physical features between prime and target were compared with those face pairs in which there was no change in physical features (see Table 1).

Avatar trials

The contrast 'within plus between versus same' was applied to face pairs in which an avatar was the prime, and this revealed right-lateralised increased activations in the fusiform gyrus (BA37), superior temporal gyrus (BA 22), and middle frontal gyrus (BA6). No differences between short and long ISI were found.

Human trials

Similarly, the contrast 'within plus between versus same' for face pairs in which a human was the prime revealed modulation of activation bilaterally in the fusiform gyrus (BA37), left precuneus (BA4), right mid-cingulum (BA 32), and left insula (BA13). No differences between short and long ISI were found.

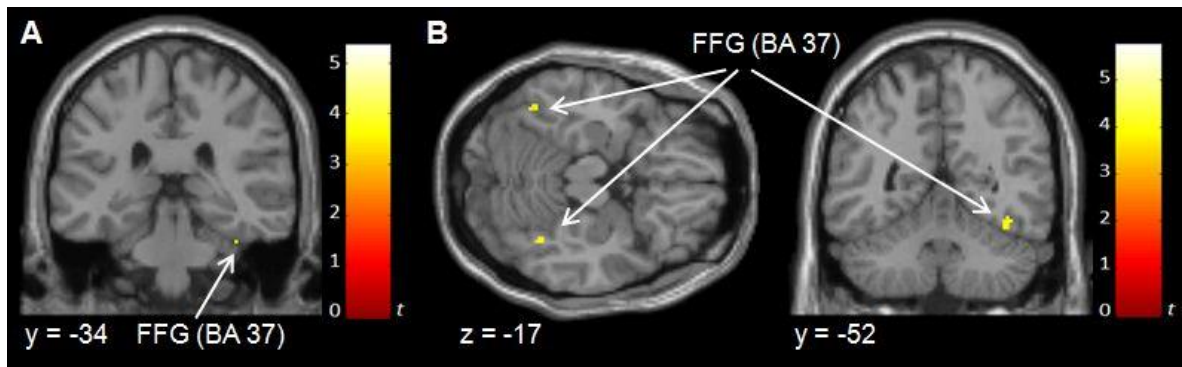


Figure 4. Effects of physical change on hemodynamic activity during observation of avatar (A) and human (B) trials.

Sensitivity to category change

To detect brain regions selectively responsive to category change across the boundary in the avatar-to-human and human-to-avatar directions along the continuum, face pairs in which there was a change in category between prime and target were compared with the conditions in which there was no such change in category (see Table 2).

Avatar-to-human direction along DOH

For the avatar-to-human direction, the contrast 'between versus same plus within' was applied to face pairs in which an avatar was the prime. This showed (at $p < .005$) a large cluster that included the left hippocampus (BA 28), entorhinal area (BA34), perirhinal area (BA35), and amygdala, and a further cluster in the right mid-insula (BA13). The activation in the left hippocampus (BA 28) was found also at the higher significance threshold of $p < .001$ (max t value 4.12 ($k = 12$) at -24, -14, -14). No differences between short and long ISI were found.

<i>Region of activation</i>	<i>BA</i>	<i>L/R</i>	<i>x</i>	<i>y</i>	<i>z</i>	<i>Max t value</i>	<i>No. of voxels</i>
<i>Change in physical qualities vs. no change in physical qualities:</i>							
Avatar							
Fusiform gyrus	37	R	38	-36	-30	4.01	8
Superior temporal gyrus	22	R	50	-60	14	4.85	16
Middle frontal gyrus	6	R	36	-2	42	4.24	6
Human							
Cerebellum		R	42	-58	-26	4.01	31
° Fusiform gyrus	37	R	42	-52	-18	3.84	
Fusiform gyrus	37	L	-46	-58	-18	3.75	5
Precuneus	4	L	-14	-42	58	5.73	33
Mid-cingulate Cortex	32	R	10	18	38	4.70	31
Insula	13	L	-38	20	8	4.01	5

° = subpeaks of a cluster; BA = approximate Brodmann area; L = left hemisphere; R = right hemisphere; MNI voxel coordinates; listed brain regions exceeding p (uncorr.) $< .001$

Table 1. Brain areas sensitive to physical change.

Human-to-avatar direction along DOH

To establish the pattern of activation associated with brain regions sensitive to crossing the category boundary in the direction of human-to-avatar, the same contrast ('between versus same plus within') was applied for face pairs in which a human was the prime. This showed a different pattern of brain areas sensitive to category change (at $p < .005$), including left and right putamen, left caudate head, right lateral posterior nucleus of thalamus, and left red nucleus. Both the right thalamus (lateral posterior nucleus) (max t value 3.88 ($k = 13$) at 14, -22, 12), left caudate head (max t value 3.62 ($k = 5$) at 12, 14, -6) and left putamen (max t value 3.88 ($k = 5$) at -20, 6, 14) were found also at the higher significance threshold of $p < .001$. No differences between short and long ISI were found.

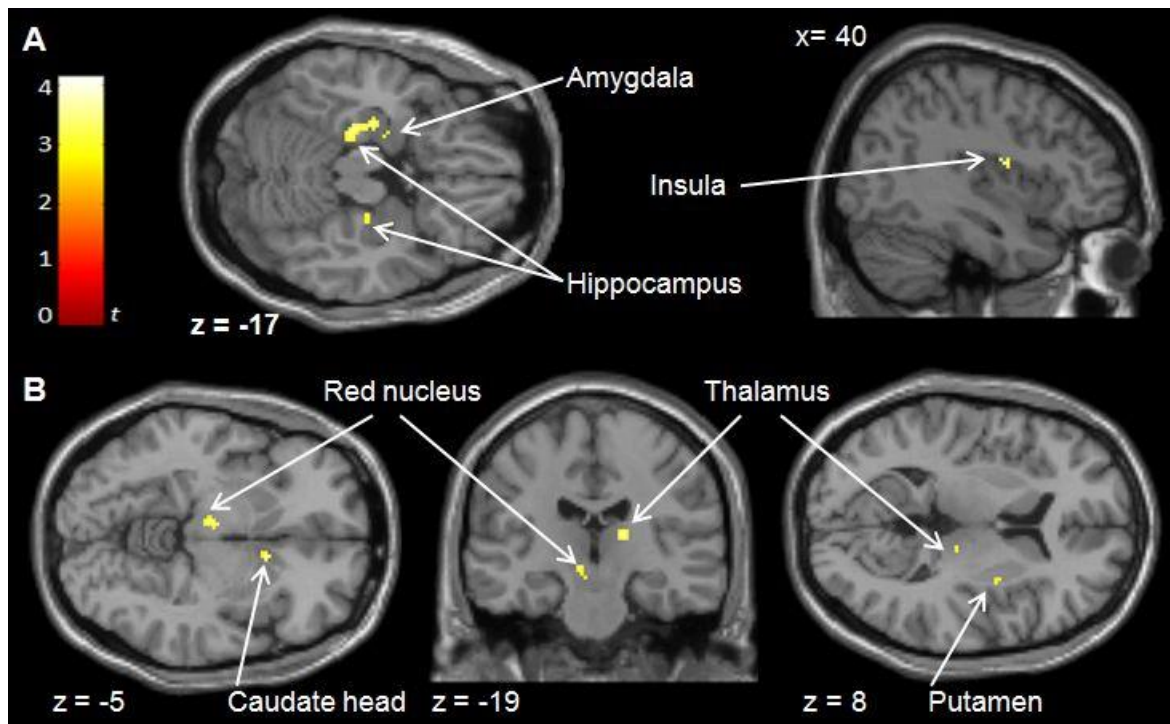


Figure 5. Effects of category change in the avatar-to-human direction (A) and human-to-avatar direction (B) along continua of human likeness.

Region of activation	BA	L/R	x	y	z	Max t value	No. of voxels
<i>Change in physical qualities vs. no change in physical qualities:</i>							
Avatar							
Fusiform gyrus	37	R	38	-36	-30	4.01	8
Superior temporal gyrus	22	R	50	-60	14	4.85	16
Middle frontal gyrus	6	R	36	-2	42	4.24	6
Human							
Cerebellum		R	42	-58	-26	4.01	31
° Fusiform gyrus	37	R	42	-52	-18	3.84	
Fusiform gyrus	37	L	-46	-58	-18	3.75	5
Precuneus	4	L	-14	-42	58	5.73	33
Mid-cingulate Cortex	32	R	10	18	38	4.70	31
Insula	13	L	-38	20	8	4.01	5

° = subpeaks of a cluster; BA = approximate Brodmann area; L = left hemisphere; R = right hemisphere; MNI voxel coordinates; listed brain regions exceeding p (uncorr.) $< .001$

Table 2. Brain areas sensitive to category change.

3.8. Discussion

The forced-choice classification task showed that faces drawn from continua representing the DOH are subjectively assigned to the discrete avatar and human categories. Because morphed faces are computer-generated and modelled, Seyama and Nagayama (2010) suggest that they are artificial and might therefore be processed differently than are natural human faces. But this view does not apply for the outcome of our perceptual category judgements: Morph stimuli regarded as artificial in a technical sense can be explicitly judged to be human. Importantly, the faces assigned to the human category were drawn from different points along the morph continua, thus demonstrating that there is variation in human-like appearance within the human category. Mori did not consider this in his hypothesis, focussing instead on the wide variation of potential physical forms of nonhuman objects along the DOH as often found in uncanny research (e.g. Walters et al., 2008; Minato et al., 2006). Ramey (2005) reflects that these objects are designed and modelled after the human image. The implication of this design and investigative approach to the uncanny is that the human image is often treated as a general point of reference irrespective of the fact, as shown in the present study, that there are differences in human likeness within the human category and its cognitive representation.

Mori's particular concern was with the threshold of nonhuman and human object perception and associated uncertainty and discomfort. Defining the nonhuman-human category boundary is therefore critical to examining Mori's hypothesis. The classification response function of all our continua showed a sigmoid shape consistent with the presence of an avatar-human category boundary (Harnad, 1987). The reaction time data indicated significant increases in and a similar degree of decision uncertainty for morphs close to each side of the category boundary. One could reframe Mori's hypothesis in terms of category processing and suggest that any personal discomfort associated with processing human likeness is most likely to occur in response to stimuli at or near the category boundary for which there is greatest categorisation ambiguity. It is possible that this prediction applies to stimuli either side of the avatar-human category boundary. This awaits investigation.

The perceptual discrimination task showed that a pair of faces located either side of the avatar-human category boundary is easier to distinguish than a pair of faces (with the same degree of physical difference along the DOH) drawn from within a category, thus indicating CP for the DOH. In other words, the cognitive representation of the category structure of the DOH influences the observer's sensitivity for perceptual information relating to human likeness. Evidently, this sensitivity is greatest at or near to the threshold of nonhuman and human object perception at which small changes in realism (i.e. concerning human or nonhuman-specifying perceptual features) might evoke uncertainty and discomfort. Taken together, these behavioural findings demonstrate that the DOH may be defined both in terms of a gradual change in the degree of physical human-like similarity, as described by Mori, as well as in terms of the effects of CP and the dimension's underlying category structure, as proposed in this investigation.

The stimulus conditions described by Mori (i.e. observing novel nonhuman objects subtly different in physical appearance from that of the human counterpart) were simulated within the constraints of fMRI methodology. The imaging data confirmed that different regions of the brain are responsive to processing change in physical human-like similarity and processing change in category along the DOH. We consider the findings for physical change first. The right and left mid-fusiform areas were sensitive to fine-grained change in physical human likeness for human trials and a different right mid-fusiform area was sensitive to physical change for avatar trials. This is consistent with reports in other studies of sensitivity in these areas to facial physical similarity (Xu et al., 2009), to similarity of facial geometry (Jiang et al., 2009; Jiang et al., 2006) and to similarity of surface properties of facial texture (e.g. Jiang et al., 2006). Our fMRI data give no indication as to the relative importance of shape and texture in processing faces along the DOH (for face recognition see, e.g. Russell et al., 2007; Jiang et al., 2006; O'Toole et al., 1999). But participants consistently reported attending to the skin texture of avatars in the forced choice classification task. Surface cues are an important diagnostic aid for judging human likeness (MacDorman et al., 2009a) and for distinguishing synthetic and natural objects and faces (e.g. Biederman and Ju, 1998; Russell and Sinha, 2007), especially when the objects are highly similar in structure (Price & Humphreys, 1989). The reported attention to skin texture may reflect greater difficulty with or preferential

use of specific facial information for forced choice classification of avatars, but this is not clear. As performance in visual discriminations of facial surface properties and shape does depend on experience (Balas & Nelson, 2010; Vuong et al., 2005), there should at least be differences between the processing of perceptual features of human and novel avatar faces. An additional analysis comparing the avatar and human "within" conditions in our perceptual discrimination task supports this: Fine visual discrimination of avatars was less accurate than that of human faces, $F(1,19) = 6,31$, $p = .02$. Whether and how the relative importance of structural and textural information changes along the course of the DOH has not been investigated (but see MacDorman et al., 2009).

A change in the category of sequentially presented faces (i.e. prime and target of paired faces) was found to evoke modulations of neural activity in brain regions previously associated with category learning and category uncertainty. The pattern of regions was entirely different depending on whether the target was an avatar in the human-avatar pairs or human in the avatar-human pairs. This shows that the direction of change along the DOH does influence the way in which humanlike faces are processed, at least within our paradigm. The dorsal and ventral striatum (putamen and head of caudate), thalamus, and red nucleus were responsive to the avatar target and the medial temporal lobe (MTL) (hippocampus, entorhinal and perirhinal areas), mid-insula and amygdala were responsive to the human target.

The results for the striatum and MTL were the most prominent. These regions are thought to have different roles in and an antagonistic relationship during category processing and learning (e.g. Seger & Cincotta, 2005; for overview see Poldrack & Rodriguez, 2004). For example, relative activity in the MTL memory system subsides and that of the striatal memory system increases during categorisation training (Poldrack et al., 1999, Poldrack et al., 2001). Alternatively, these and other memory systems (e.g. Poldrack & Foerde, 2008) might actually work in concert during category processing (Cincotta & Seger, 2007). Irrespective of their possible interplay, the differential response of these regions to the direction of category change between prime and target (i.e. human-avatar or avatar-human) suggests that the avatar and human faces represent different categorisation problems that require, at least in part, dissimilar processes or strategies to resolve them. There is some support for this in

the behavioural data from the perceptual discrimination task. An additional analysis compared the avatar and human "between" conditions and found a significantly higher proportion of judgement errors [$F(1,19) = 8,69, p < .01$] for avatar-human pairs than for human-avatar pairs.

In the pair-repetition paradigm, processing of the target stimulus is biased by the implicit memory of the preceding prime that represents a different "expected" or "predicted" category than is actually shown in the target. The caudate head (responsive to avatar targets) is sensitive to prediction error in various tasks (e.g. O'Doherty et al., 2003; Bray and O'Doherty, 2007; King-Casas et al., 2005; Jensen et al., 2007). Besides contributing to learning stimulus–category associations (Seger and Cincotta, 2005, Seger and Cincotta, 2006) and selecting and executing motor responses for signalling category membership (Seger et al., 2010, Williams and Eskandar, 2006), the putamen (also responsive to avatar targets) is associated with processing prediction error (den Ouden et al., 2010). It is possible that these striatal regions contribute to processing the deviation of the novel avatar target from the human prime and do so on the basis of the prime and well defined category representations of human faces. If this is true, one would expect the "between" condition of the perceptual discrimination task to be easier for the human-avatar pairs than for the avatar-human pairs for which corresponding category representations are less well defined. The preceding additional analysis of the "between" condition supports this. Predictive processing has been investigated explicitly under conditions of decision uncertainty and shown to modulate activity in the ventral striatum (and the thalamus and red nucleus as also found in response to our targets of the avatar-human face pairs) (Volz et al., 2003). Filoteo et al. (2005) suggest that the caudate head might be involved in switching between potential category rules used to establish category membership, while Grinband et al. (2006) found activations highly similar to ours in their study of categorisation ambiguity, and they suggest that the ventral striatum signals adjustment of the represented categorical boundary in order to minimise errors.

MTL structures (responsive to human targets) are involved in processing novel and familiar stimuli, novelty detection, and uncertainty coding (e.g., Rutishauser et al., 2006; Stark and Squire, 2000; Stern et al., 1996; Vanni-Mercier et al., 2009). MTL is

thought to encode and retain individual category instances of novel stimuli (Poldrack et al., 1999, Seger and Cincotta, 2005) and to facilitate further category-related processing of features by other systems including the basal ganglia (Meeter et al., 2008). Given our paired presentation of similar human and novel avatar faces, it is noteworthy that processes of novelty detection in the hippocampus are considered to entail processing of the deviation of actual sensory input from expected input (e.g., Kumaran and Maguire, 2006). This would be consistent with the notion that the processing of our human targets was strongly guided by representations of the preceding input from the avatar primes. The hippocampus is involved in visual categorisation and perceptual learning, while impaired processing of faces is found only when both hippocampus and perirhinal cortex are damaged (Graham et al., 2006). The perirhinal cortex played a clear role in response to category change between human primes and avatar targets. The perirhinal cortex appears to be critical for object memory, contributing to resolving complex visual discriminations and those of high ambiguity (Bussey et al., 2003) when both novel and familiar stimuli share visual features (Barense et al., 2005).

Consistent with its role in decision-making under conditions of category processing and uncertainty (e.g. Fleming et al., 2010; Heekeren et al., 2008; Grinband et al., 2006; Hsu et al., 2005; Volz et al., 2003), we expected the anterior rather than the mid-insular cortex to be responsive to category change. The mid-insular cortex was responsive to human targets. The role of this region in the present study and generally (Wager and Feldman Barrett, 2004) is not clear. The mid-insular is responsive to affective ambiguity, anticipation of emotionally aversive visual stimuli, and to physical pain (Simmons et al., 2008), this reflecting the processing of negative expectation particularly in the context of pain (e.g. Petrovic et al., 2000). Our participants were not exposed to noxious stimulation, but similar neural mechanisms are thought to mediate the experience of pain and feelings of personal discomfort (Price, 2002). Right mid- (Lutz et al., 2009) and anterior insula activity (Critchley et al., 2002) is also associated with measures of arousal (heart rate and galvanic skin responses). Given the suggested role of prediction error processing, Paulus and Stein's (2006) proposed anxiety mechanism is interesting. Applied here, the deviation between the expected arousal state on the basis of the avatar prime and the actual arousal associated with the unexpected human target might be interpreted as

signalling the presence of uncertainty, threat or potential threat. Uncanny-related arousal and discomfort is suggested to be rooted in mechanisms of threat avoidant behaviour (Green et al., 2008; MacDorman, 2005a). Interestingly, Gray and Critchley (2007) and Wager et al. (2003) associate mid-insular representations with a threat-related component and avoidant behaviour. Alternatively, arousal and mid-insula responsiveness to category change may be more closely related to the role of the right anterior insula during category processing in marshalling attentional resources to enhance performance during categorisation under conditions of uncertainty (Heekeren et al., 2008).

In exploring additional brain areas otherwise associated with affective processing, the amygdala was found to be responsive to category change in avatar-human face pairs. The amygdala is responsive to natural and computer-generated human faces (e.g. Todorov and Engell, 2008), human-like but unnatural faces (Rotshtein et al., 2001) novelty, uncertainty, unclear predictive value, and ambiguous valence (e.g. Neta and Whalen, 2010; Herwig et al., 2007; Levy et al., 2010; Phelps and LeDoux, 2005). Whether and how appraisal of affective valence (and prospective outcomes in terms of potential threat or reward) contributes to category processing is not clear, though prospective outcomes are thought to interact with and bias processes of perceptual categorisation (Heekeren et al., 2008; see also Gupta et al., 2011). In view of the different responses of the striatum and MTL in this study, Seger and Miller's (2010) suggestion is interesting. They propose that the amygdala might play a role during category processing in altering the balance between the memory systems of these two regions depending on the affective meaning of a situation. For example, modulation of the amygdala might reflect its involvement in processing of novel category information in conjunction with the MTL and enhancement of attentional and memory processing of the novel avatar primes (Kleinhans et al., 2007). Alternatively, the amygdala might have been responsive to the "unexpected" category of the targets (i.e. the human faces) (Roesch et al., 2010).

This investigation demonstrates that the definition of the DOH in terms of the degree of object similarity to human appearance does not reflect the way in which human likeness is subjectively perceived along this dimension. The forced-choice categorisation task showed that there is variation in human likeness within the human

category, and the perceptual discrimination task confirmed that processes of CP and associated representations of the DOH's category structure influence the perception of human likeness. Given the inconsistent findings in "uncanny" research to date, careful definition of the category boundary especially in studies using morph continua is therefore important. It is likely that mechanisms associated with category processing and the representation of human likeness also influence affective experience. For example, increased uncertainty in forced-choice category decisions might be associated with discomfort (assuming that Mori's hypothesis is correct), though any discomfort need not be limited to the nonhuman side of the category boundary along the DOH. The present behavioural findings suggest that the general conception of the DOH should be revised. As Mori did not embed his ideas in any psychological framework, we suggest that his hypothesis should be considered in terms of the well-established psychological empirical-theoretical framework of category perception and learning. This could prove useful in examining where along the DOH negative or uncanny experience is most likely to occur for a given set of stimuli and in throwing light on the potential role of categorisation ambiguity in evoking negative affect. It is of course possible that any association between category ambiguity and negative affect is not specific to humanlike stimuli along the DOH but is a more general feature of category ambiguity itself.

The framework set out in this study could also be applied to other fields of related research to understand how variously realistic humanlike characters in association with category ambiguity influence for example the experience of presence (i.e., immersive experience) in virtual reality (e.g., Brenton et al., 2008), audience persuasion and identification with fictive characters in narrative (e.g., Cohen, 2001), communication in educational virtual environments (e.g., Fabri et al., 2004), or consumer trust in electronic commerce (e.g., Bauer et al., 2005). For example, greater category ambiguity might enhance attentional processing of the human or non-human specifying perceptual features of digital human representations, and this might render it less likely that a person responds to the character as if it were in some way real, experiences identification, processes the media content as the designers intend, or develops a sense of trust. Using this framework, the fMRI study showed that different brain regions are responsive to the direction along the DOH in which there is a change in the category of observed objects (i.e. avatar-to-human or

human-to-avatar). This appears to reflect the impact of differences in category knowledge for avatar and human faces. Re-examination of this effect in the context of category training and experience with avatars might throw more light on this. Replicating the behavioural findings of this study with a larger number of participants and with stimuli presenting biological motion (see, Saygin et al., 2011; Chaminade et al., 2007) would reinforce our findings. Whether there are gender differences remains to be investigated.

Conflict of Interest Statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgments

This work was supported by the European Union FET Integrated Project PRESENCCIA (Contract number 27731).

3.9. References

- Angeli, A., Davidoff, J., & Valentine, T. (2008). Face familiarity, distinctiveness, and categorical perception. *Q. J. Exp. Psychol. (Hove)* 61, 690–707.
- Annett, M. (1970). A classification of hand preference by association analysis. *Br. J. Psychol.* 61, 303–321.
- Ashby, F. G., & Maddox, W. T. (2005). Human category learning. *Annu. Rev. Psychol.* 56, 149–178.
- Aylett, R. S. (2004). Agents and affect: why embodied agents need affective systems. *Lect. Notes Comput. Sci.* 3025, 496–504.
- Balas, B., & Nelson, C. A. (2010). The role of face shape and pigmentation in other-race face perception: an electrophysiological study. *Neuropsychologia* 48, 498–506.
- Barense, M., Bussey, T., Lee, A., Rogers, T. T., Hodges, J. R., Saksida, L., Murray, E., & Graham, K. (2005). Feature ambiguity influences performance on novel object discriminations in patients with damage to perirhinal cortex. *Cognitive Neuroscience Society Annual Meeting Program 2005*, New York, NY, 125–116.
- Bauer, H. H., & Neumann, M. (2005). “Investigating the Effects of Avatars as Virtual Representatives in Electronic Commerce,” in *Proceedings of the Australian-New Zealand Marketing Academy Conference (ANZMAC)*, Perth, 18.

- Biederman, I., & Ju, G. (1998). Surface vs. edge-based determinants of visual recognition. *Cogn. Psychol.* 20, 38–64.
- Bray, S., & O'Doherty, J. (2007). Neural coding of reward-prediction error signals during classical conditioning with attractive faces. *J. Neurophysiol.* 97, 3036–3045.
- Brenton, H., Gillies, M., Ballin, D., & Chatting, D. (2008). The uncanny valley: does it exist? *Paper presented at the Proceedings of the Human-Animated Characters Interaction, HCI 2005: The Bigger Picture*, Edinburgh.
- Bussey, T. J., Saksida, L. M., & Murray, E. A. (2003). Impairments in visual discrimination after perirhinal cortex lesions: testing “declarative” versus “perceptual-mnemonic” views of perirhinal cortex function. *Eur. J. Neurosci.* 17, 649–660.
- Campbell, R., Pascalis, O., Coleman, M., Wallace, S. B., & Benson, P. J. (1987). Are faces of different species perceived categorically by human observers? *Proc. R. Soc. Lond. B Biol. Sci.* 264, 1429–1434.
- Chaminade, T., Hodgins, J., & Kawato, M. (2007). Anthropomorphism influences perception of computer-animated characters' actions. *Soc. Cogn. Affect. Neurosci.* 2, 206–216.
- Cincotta, C. M., & Seger, C. A. (2007). Dissociation between striatal regions while learning to categorize via feedback and via observation. *J. Cogn. Neurosci.* 19, 249–265.
- Cohen, J. (2001). Defining identification: a theoretical look at the identification of audiences with media characters. *Mass Commun. Soc.* 4, 245–264.
- Critchley, H. D., Mathias, C. J., & Dolan, R. J. (2002). Fear conditioning in humans: the influence of awareness and autonomic arousal on functional neuroanatomy. *Neuron* 33, 653–663.
- den Ouden, H. E., Daunizeau, J., Roiser, J., Friston, K. J., & Stephan, K. E. (2010). Striatal prediction error modulates cortical coupling. *J. Neurosci.* 30, 3210–3219.
- Etkin, A., Egner, T., Peraza, D. M., Kandel, E. R., and Hirsch, J. (2006). Resolving emotional conflict: a role for the rostral anterior cingulate cortex in modulating activity in the amygdala. *Neuron* 51, 871–882.
- Fabri, M., Moor, D., & Hobbs, D. (2004). Mediating the expression of emotion in educational collaborative virtual environments: an experimental study. *Int. J. Virtual Reality* 7, 66–81.
- Filoteo, J. V., Maddox, W. T., Salmon, D. P., & Song, D. D. (2005). Information-integration category learning in patients with striatal dysfunction. *Neuropsychology* 19, 212–222.
- Fleming, S. M., Whiteley, L., Hulme, O. J., Sahani, M., and Dolan, R. J. (2010). Effects of category-specific costs on neural systems for perceptual decision-making. *J. Neurophysiol.* 103, 3238–3247.
- Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J. P., Frith, C. D., & Frackowiak, R. S. J. (1995). Statistical parametric maps in functional imaging: a general linear approach. *Hum. Brain Mapp.* 2, 189–210.
- Graham, K. S., Scahill, V. L., Hornberger, M., Barense, M. D., Lee, A. C., Bussey, T. J., & Saksida, L. M. (2006). Abnormal categorization and perceptual learning in patients with hippocampal damage. *J. Neurosci.* 26, 7547–7554.
- Gray, M. A., & Critchley, H. D. (2007). Interoceptive basis to craving. *Neuron* 54, 183–186.

- Green, R. D., MacDorman, K. F., Ho, C.-C., & Vasudevan, S. K. (2008). Sensitivity to the proportions of faces that vary in human likeness. *Comput. Hum. Behav.* 24, 2456–2474.
- Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: neural models of stimulus-specific effects. *Trends Cogn. Sci. (Regul. Ed.)* 10, 14–23.
- Grill-Spector, K., & Malach, R. (2001). fMR-adaptation: a tool for studying the functional properties of human cortical neurons. *Acta Psychol. (Amst.)* 107, 293–321.
- Grinband, J., Hirsch, J., & Ferrera, V. P. (2006). A neural representation of categorization uncertainty in the human brain. *Neuron* 49, 757–763.
- Gupta, R., Koscik, T. R., Bechara, A., & Tranel, D. (2011). The amygdala and decision-making. *Neuropsychologia* 49, 760–766.
- Hanson, D. (2006). Exploring the aesthetic range for humanoid robots. *Paper Presented at the Proceedings of the ICCS/CogSci-2006 Long Symposium: Toward Social Mechanisms of Android Science*, Vancouver.
- Harnad, S. R. (1987). “Introduction: psychological and cognitive aspects of categorical perception: a critical overview,” in *Categorical Perception: The Groundwork of Cognition*, ed. S. Harnad (New York: Cambridge University Press), 1–25.
- Heekeren, H. R., Marrett, S., & Ungerleider, L. G. (2008). The neural systems that mediate human perceptual decision making. *Nat. Rev. Neurosci.* 9, 467–479.
- Henson, R. N. (2003). Neuroimaging studies of priming. *Prog. Neurobiol.* 70, 53–81.
- Herwig, U., Kaffenberger, T., Baumgartner, T., and Jancke, L. (2007). Neural correlates of a “pessimistic” attitude when anticipating events of unknown emotional valence. *Neuroimage* 34, 848–858.
- Ho, C.-C., & MacDorman, K. (2010). Revisiting the uncanny valley theory: developing and validating an alternative to the Godspeed indices. *Comput. Hum. Behav.* 26, 1508–1518.
- Hsu, M., Bhatt, M., Adolphs, R., Tranel, D., & Camerer, C. F. (2005). Neural systems responding to degrees of uncertainty in human decision-making. *Science* 310, 1680–1683.
- Jensen, J., Smith, A. J., Willeit, M., Crawley, A. P., Mikulis, D. J., Vitcu, I., & Kapur, S. (2007). Separate brain regions code for salience vs. valence during reward prediction in humans. *Hum. Brain Mapp.* 28, 294–302.
- Jiang, F., Dricot, L., Blanz, V., Goebel, R., & Rossion, B. (2009). Neural correlates of shape and surface reflectance information in individual faces. *Neuroscience* 163, 1078–1091.
- Jiang, X., Bradley, E., Rini, R. A., Zeffiro, T., Vanmeter, J., & Riesenhuber, M. (2007). Categorization training results in shape- and category-selective human neural plasticity. *Neuron* 53, 891–903.
- Jiang, X., Rosen, E., Zeffiro, T., Vanmeter, J., Blanz, V., & Riesenhuber, M. (2006). Evaluation of a shape-based model of human face discrimination using fMRI and behavioral techniques. *Neuron* 50, 159–172.
- King-Casas, B., Tomlin, D., Anen, C., Camerer, C. F., Quartz, S. R., & Montague, P. R. (2005). Getting to know you: reputation and trust in a two-person economic exchange. *Science* 308, 78–83.
- Kleinhans, N. M., Johnson, L. C., Mahurin, R., Richards, T., Stegbauer, K. C., Greenson, J., Dawson, G., & Aylward, E. (2007). Increased amygdala activation to neutral faces is associated with better face memory performance. *Neuroreport* 18, 987–991.

- Kumaran, D., & Maguire, E. A. (2006). An unexpected sequence of events: mismatch detection in the human hippocampus. *PLoS Biol.* 4, e424. doi: 10.1371/journal.pbio.0040424
- Levy, I., Snell, J., Nelson, A. J., Rustichini, A., & Glimcher, P. W. (2010). Neural representation of subjective value under risk and ambiguity. *J. Neurophysiol.* 103, 1036–1047.
- Li, S., Mayhew, S. D., & Kourtzi, Z. (2009). Learning shapes the representation of behavioral choice in the human brain. *Neuron* 62, 441–452.
- Lutz, A., Greischar, L. L., Perlman, D. M., & Davidson, R. J. (2009). BOLD signal in insula is differentially related to cardiac function during compassion meditation in experts vs. novices. *Neuroimage* 47, 1038–1046.
- MacDorman, K. (2005a). "Mortality salience and the uncanny valley," in *IEEE-RAS International Conference on Humanoid Robots* (Tsukuba: IEEE), 339–405.
- MacDorman, K. (2005b). "Androids as an Experimental Apparatus: Why is There an Uncanny Valley and Can We Exploit it?" in *Toward Social Mechanisms of Android Science: A CogSci 2005 Workshop*, Stresa, 106–118.
- MacDorman, K. (2006). Subjective ratings of robot video clips for human likeness, familiarity, and eeriness: an exploration of the uncanny valley. *Paper presented at the ICCS/CogSci-2006 Long Symposium: Toward Social Mechanisms of Android Science*, Vancouver.
- MacDorman, K., Green, R., Ho, C.-C., & Koch, C. (2009). Too real for comfort? Uncanny responses to computer generated faces. *Comput. Hum. Behav.* 25, 695–710.
- MacDorman, K., & Ishiguro, H. (2006). The uncanny advantage of using androids in cognitive science research. *Interact. Stud.* 7, 297–337.
- Maddox, W. T., & Ashby, F. G. (2004). Dissociating explicit and procedural-learning based systems of perceptual category learning. *Behav. Processes* 66, 309–332.
- Meeter, M., Radics, G., Myers, C. E., Gluck, M. A., & Hopkins, R. O. (2008). Probabilistic categorization: how do normal participants and amnesic patients do it? *Neurosci. Biobehav. Rev.* 32, 237–248.
- Minato, T., Shimada, M., Itakura, S., Lee, K., & Ishiguro, H. (2006). Evaluating the human likeness of an android by comparing gaze behaviors elicited by the android and a person. *Adv. Robot.* 20, 1147–1163.
- Mori, M. (1970). Bukimi no tani [the uncanny valley]. *Energy* 7, 33–35.
- Murray, S. O., & Wojciulik, E. (2004). Attention increases neural selectivity in the human lateral occipital complex. *Nat. Neurosci.* 7, 70–74.
- Neta, M., & Whalen, P. J. (2010). The primacy of negative interpretations when resolving the valence of ambiguous facial expressions. *Psychol. Sci.* 21, 901–907.
- O'Doherty, J. P., Dayan, P., Friston, K., Critchley, H., & Dolan, R. J. (2003). Temporal difference models and reward-related learning in the human brain. *Neuron* 38, 329–337.
- O'Toole, A. J., Vetter, T., & Blanz, V. (1999). Three-dimensional shape and two-dimensional surface reflectance contributions to face recognition: an application of three-dimensional morphing. *Vision Res.* 39, 3145–3155.
- Pastore, R. E. (1987). "Categorical perception: some psychophysical models," in *Categorical perception: The groundwork of Cognition*, ed. S. Harnad (New York: Cambridge University Press), 29–52.
- Paulus, M. P., & Stein, M. B. (2006). An insular view of anxiety. *Biol. Psychiatry* 60, 383–387.

- Petrovic, P., Petersson, K. M., Ghatan, P. H., Stone-Elander, S., & Ingvar, M. (2000). Pain-related cerebral activation is altered by a distracting cognitive task. *Pain* 85, 19–30.
- Phelps, E. A. (2004). Human emotion and memory: Interactions of the amygdala and hippocampal complex. *Curr. Opin. Neurobiol.* 14, 198–202.
- Phelps, E. A., & LeDoux, J. E. (2005). Contributions of the amygdala to emotion processing: from animal models to human behavior. *Neuron* 48, 175–187.
- Phelps, E. A., O'Connor, K. J., Gatenby, J. C., Gore, J. C., Grillon, C., & Davis, M. (2008). Activation of the left amygdala to a cognitive representation of fear. *Neurosciences* 4, 37–41.
- Poldrack, R. A., Clark, J., Pare-Blagoev, E. J., Shohamy, D., Creso Moyano, J., Myers, C., & Gluck, M. A. (2001). Interactive memory systems in the human brain. *Nature* 414, 546–550.
- Poldrack, R. A., & Foerde, K. (2008). Category learning and the memory systems debate. *Neurosci. Biobehav. Rev.* 32, 197–205.
- Poldrack, R. A., Prabhakaran, V., Seger, C. A., & Gabrieli, J. D. (1999). Striatal activation during acquisition of a cognitive skill. *Neuropsychology* 13, 564–574.
- Poldrack, R. A., & Rodriguez, P. (2004). How do memory systems interact? Evidence from human classification learning. *Neurobiol. Learn. Mem.* 82, 324–332.
- Pollick, F. (2009). “In search of the Uncanny Valley,” in *Lecture notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, UC Media 2009*, eds P. Daras, and O. M. Ibarra (Venice: Springer), 69–78.
- Price, D. D. (2000). Psychological and neural mechanisms of the affective dimension of pain. *Science* 288, 1769–1772.
- Price, C. J., & Humphreys, G. W. (1989). The effects of surface detail on object categorization and naming. *Q. J. Exp. Psychol. A.* 41, 797–827.
- Pruessmann, K. P., Weiger, M., Scheidegger, M. B., & Boesinger, P. (1999). SENSE: Sensitivity encoding for fast MRI. *Magn. Reson. Med.* 42, 952–962.
- Ramey, C. H. (2005). The uncanny valley of similarities concerning abortion, baldness, heaps of sand, and humanlike robots. *Paper Presented at the Proceedings of Views of the Uncanny Valley Workshop: IEEE-RAS International Conference on Humanoid Robots*, Tsukuba.
- Roesch, M. R., Calu, D. J., Esber, G. R., & Schoenbaum, G. (2010). All that glitters. Dissociating attention and outcome expectancy from prediction errors signals. *J. Neurophysiol.* 104, 587–595.
- Rotshtein, P., Henson, R. N., Treves, A., Driver, J., & Dolan, R. J. (2005). Morphing Marilyn into Maggie dissociates physical and identity face representations in the brain. *Nat. Neurosci.* 8, 107–113.
- Rotshtein, P., Malach, R., Hadar, U., Graif, M., & Hendler, T. (2001). Feeling or features: different sensitivity to emotion in high-order visual cortex and amygdala. *Neuron* 32, 747–757.
- Russell, R., Biederman, I., Nederhouser, M., & Sinha, P. (2007). The utility of surface reflectance for the recognition of upright and inverted faces. *Vision Res.* 47, 157–165.
- Russell, R., and Sinha, P. (2007). Real-world face recognition: the importance of surface reflectance properties. *Perception* 36, 1368–1374.
- Rutishauser, U., Mamelak, A. N., & Schuman, E. M. (2006). Single-trial learning of novel stimuli by individual neurons of the human hippocampus-amygdala complex. *Neuron* 49, 805–813.

- Saygin, A. P., Chaminade, T., Ishiguro, H., Driver, J., & Frith, C. (2011). The thing that should not be: predictive coding and the uncanny valley in perceiving human and humanoid robot actions. *Soc. Cogn. Affect. Neurosci.* doi: 10.1093/scan/nsr025
- Schultz, J., & Pilz, K. S. (2009). Natural facial motion enhances the cortical responses to faces. *Exp. Brain Res.* 194, 465–475.
- Seger, C. A., & Cincotta, C. M. (2005). The roles of the caudate nucleus in human classification learning. *J. Neurosci.* 25, 2941–2951.
- Seger, C. A., & Cincotta, C. M. (2006). Dynamics of frontal, striatal, and hippocampal systems during rule learning. *Cereb. Cortex* 16, 1546–1555.
- Seger, C. A., & Miller, E. K. (2010). Category learning in the brain. *Annu. Rev. Neurosci.* 33, 203–219.
- Seger, C. A., Peterson, E. J., Cincotta, C. M., Lopez-Paniagua, D., & Anderson, C. W. (2010). Dissociating the contributions of independent corticostriatal systems to visual categorization learning through the use of reinforcement learning modeling and Granger causality modeling. *Neuroimage* 50, 644–656.
- Seyama, J., & Nagayama, R. S. (2007). The uncanny valley: effect of realism on the impression of artificial human faces. *Presence (Camb.)* 16, 337–351.
- Seyama, J., & Nagayama, R. S. (2009). Probing the uncanny valley with the eye size aftereffect. *Presence-Teleop. Virt.* 18, 321–339.
- Simmons, A., Matthews, S. C., Paulus, M. P., & Stein, M. B. (2008). Intolerance of uncertainty correlates with insula activation during affective ambiguity. *Neurosci. Lett.* 430, 92–97.
- Stark, C. E., & Squire, L. R. (2000). Functional magnetic resonance imaging (fMRI) activity in the hippocampal region during recognition memory. *J. Neurosci.* 20, 7776–7781.
- Stern, C. E., Corkin, S., Gonzalez, R. G., Guimaraes, A. R., Baker, J. R., Jennings, P. J., Carr, C. A., Sugiura, R. M., Vedantham, V., & Rosen, B. R. (1996). The hippocampal formation participates in novel picture encoding: evidence from functional magnetic resonance imaging. *Proc. Natl. Acad. Sci. U.S.A.* 93, 8660–8665.
- Studdert-Kennedy, M., Lineman, A. M., Hassis, K. S., & Cooper, F. R. (1970). Motor theory of speech perception: a reply to Lane's critical review. *Psychol. Rev.* 77, 234–249.
- Tinwell, A., & Grimshaw, M. (2009). "Bridging the uncanny: an impossible traverse?" in *Proceedings of the 13th International MindTrek Conference: Everyday Life in the Ubiquitous Era*, eds. O. Sotamaa, A. Lugmayr, H. Franssila, P. Näränen, and J. Vanhala (Tampere: ACM), 66–73.
- Tinwell, A., Grimshaw, M., & Williams, A. (2010). Uncanny behaviour in survival horror games. *J. Gaming Virtual Worlds* 2, 3–25.
- Todorov, A., & Engell, A. (2008). The role of the amygdala in implicit evaluation of emotionally neutral faces. *Soc. Cogn. Affect. Neurosci.* 3, 303–312.
- Vanni-Mercier, G., Mauguiere, F., Isnard, J., & Dreher, J. C. (2009). The hippocampus codes the uncertainty of cue-outcome associations: an intracranial electrophysiological study in humans. *J. Neurosci.* 29, 5287–5294.
- Volz, K. G., Schubotz, R. I., & von Cramon, D. Y. (2003). Predicting events of varying probability: uncertainty investigated by fMRI. *Neuroimage* 19(Pt 1), 271–280.
- Vuilleumier, P. (2005). How brains beware: neural mechanisms of emotional attention. *Trends Cogn. Sci. (Regul. Ed.)* 9, 585–594.

- Vuong, Q. C., Peissig, J. J., Harrison, M. C., & Tarr, M. J. (2005). The role of surface pigmentation for recognition revealed by contrast reversal in faces and Greebles. *Vision Res.* 45, 1213–1223.
- Wager, T. D., & Feldman Barrett, L. (2004). From affect to control: functional specialization of the insula in motivation and regulation. Available at: http://www.affective-science.org/pubs/2004/Wager_Edfest_submitted_copy.pdf [Published online at PsycExtra].
- Wager, T. D., Phan, K. L., Liberzon, I., & Taylor, S. F. (2003). Valence, gender, and lateralization of functional brain anatomy in emotion: a meta-analysis of findings from neuroimaging. *Neuroimage* 19, 513–531.
- Wages, R., Grünvogel, S. M., & Grützmacher, B. (2004). How realistic is realism? Considerations on the aesthetics of computer games. *Lect. Notes Comput. Sci.* 3166, 216–225.
- Walters, M. L., Syrdal, D. S., Dautenhaun, K., te Boekhorst, R., & Koay, K. L. (2008). Avoiding the uncanny valley: robot appearance, personality and consistency of behavior in a an attention-seeking home scenario for a robot companion. *Auton. Robots* 24, 159–178.
- Weiskopf, N., Hutton, C., Joseph, O., Turner, R., & Deichmann, R. (2007). Optimized EPI for fMRI studies of the orbitofrontal cortex: compensation of susceptibility-induced gradients in the readout direction. *MAGMA* 20, 39–49.
- Williams, Z. M., & Eskandar, E. N. (2006). Selective enhancement of associative learning by microstimulation of the anterior caudate. *Nat. Neurosci.* 9, 562–568.
- Xu, X., Yue, X., Lescroart, M. D., Biederman, I., & Kim, J. G. (2009). Adaptation in the fusiform face area (FFA): image or person? *Vision Res.* 49, 2800–2807.
- Yamamoto, K., Tanaka, S., Kobayashi, H., Kozima, H., & Hashiya, K. (2009). A non-humanoid robot in the “uncanny valley”: experimental analysis of the reaction to behavioral contingency in 2-3 year old children. *PLoS ONE* 4, e6974. doi: 10.1371/journal.pone.0006974

4. Study B. Category processing and the human likeness dimension of the uncanny valley hypothesis: eye-tracking data

4.1. Abstract

The Uncanny Valley Hypothesis (Mori, 1970) predicts that perceptual difficulty distinguishing between a humanlike object (e.g., lifelike prosthetic hand, mannequin) and its human counterpart evokes negative affect. Research has focused on affect, with inconsistent results, but little is known about how objects along the hypothesis' dimension of human likeness (DHL) are actually perceived. This study used morph continua based on human and highly realistic computer-generated (avatar) faces to represent the DHL. Total number and dwell time of fixations to facial features were recorded while participants ($N = 60$) judged avatar versus human category membership of the faces in a forced choice categorization task. Fixation and dwell data confirmed the face feature hierarchy (eyes, nose, and mouth in this order of importance) across the DHL. There were no further findings for fixation. A change in the relative importance of these features was found for dwell time, with greater preferential processing of eyes and mouth of categorically ambiguous faces compared with unambiguous avatar faces. There were no significant differences between ambiguous and human faces. These findings applied for men and women, though women generally dwelled more on the eyes to the disadvantage of the nose. The mouth was unaffected by gender. In summary, the relative importance of facial features changed on the DHL's non-human side as a function of categorization ambiguity. This change was indicated by dwell time only, suggesting greater depth of perceptual processing of the eyes and mouth of ambiguous faces compared with these features in unambiguous avatar faces.

4.2. Introduction

The Uncanny Valley Hypothesis (Mori, 1970) predicts that difficulty distinguishing between a humanlike object (e.g., robot, lifelike prosthetic hand, mannequin) and its natural human counterpart will evoke negatively valenced feelings and cognitions (for recent overviews, see, MacDorman et al., 2009a; Pollick, 2010). Mori described this negative state as characterized by feelings of unease and the uncanny (Figure 1).

This prediction has been of concern to animators, video game designers and roboticists in their effort to ensure that the appearance of highly realistic humanlike characters influences subjective experience and behaviour in the way that its design intended (e.g., Fabri et al., 2004; Minato et al., 2006; Walters et al., 2008; MacDorman et al., 2009a; Ho and MacDorman, 2010). Uncanny experience and the hypothesis' valence dimension have therefore received much research attention. But this research has produced inconsistent results (e.g., Hanson, 2006; MacDorman, 2006; Tinwell and Grimshaw, 2009; Tinwell et al., 2010). One likely reason for this inconsistency is the uncertainty surrounding the vague terminology used to describe the hypothesis' valence dimension (e.g., Bartneck et al., 2007; Seyama and Nagayama, 2007; MacDorman et al., 2009a; Dill et al., 2012). Another reason might be related to the fact that the hypothesis' dimension of human likeness (DHL), defined in terms of a linear change in the degree of physical humanlike similarity, is not subjectively perceived as a simple linear change in humanlike similarity (Cheetham et al., 2011; Cheetham and Jancke, 2013). Given the aim of uncanny research to understand subjective experience of nonhuman characters in terms of human likeness, this recent evidence suggests that a better understanding of how human likeness is really perceived is needed.

One approach to investigating perception along the DHL is to record eye movement (Just and Carpenter, 1980; Goldberg and Wichansky, 2003). Eye movements orient foveal vision to perceptual details that are critical for encoding and perceiving objects (Yarbus, 1967; Loftus and Mackworth, 1978; Desimone and Duncan, 1995; Egeth and Yantis, 1997 (for reviews, see Rayner, 1998; Henderson, 2003, 2007). Investigation of eye movement has contributed to our understanding of face perception (e.g., Walker-Smith et al., 1977; Cook, 1978; Schyns et al., 2002; Pearson et al., 2003; Stacey et al., 2005). These studies show that measures of eye movement such as the number of fixations and dwell time (i.e., the cumulative duration of fixations) over a particular area of interest are sensitive to bottom-up effects of stimulus information and to top-down task requirements (e.g., Robinson, 1964; Yarbus, 1967; Posner, 1980).

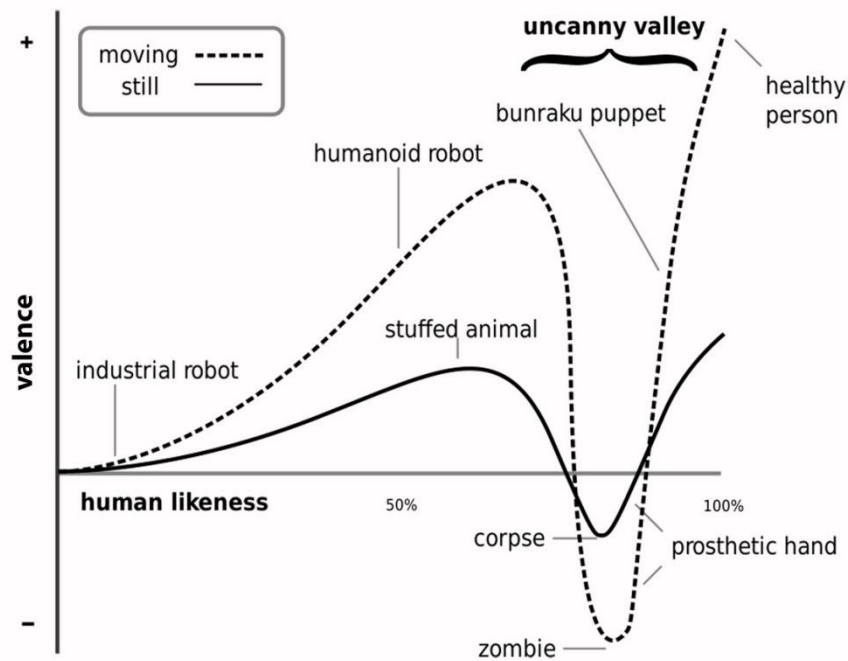


Figure 1. Uncanny valley hypothesis. The illustration shows the hypothesis' proposed non-linear relationship between physical humanlike realism and affective experience. This relationship is generally positively valenced except at the degree of humanlike realism at which an object's appearance and behavior is difficult to distinguish from that of the real human counterpart (i.e., at the first positive peak). This difficulty results in a sharp negative peak in the valence of affective experience that Mori describes as including feelings of eeriness and the uncanny (i.e., uncanny valley). This uncanny effect is thought to be stronger for objects in motion (illustration adapted from MacDorman, 2005).

The Uncanny Valley Hypothesis essentially describes a situation in which individuals engage in the task of implicitly or explicitly assigning perceptually similar humanlike objects to a non-human or human category. An implicit assumption of the hypothesis is that this task can vary in difficulty along the DHL. This was confirmed by Cheetham et al. (2011). They used human and highly realistic avatar faces to generate morph continua to represent the DHL and presented the morphs of these continua in a two-alternative forced choice categorization task. This task required participants to assign membership of each morphed face to either the human or avatar category. Their measures of response accuracy and response time confirmed that task difficulty increases sharply at the category boundary, that is, at the point of greatest categorization ambiguity. Similar findings have been reported for continua based on

other stimuli (e.g., Campbell et al., 1997). Consistent with Mori's informal description of the hypothesis, this boundary should therefore mark the point along the DHL at which perceptual difficulty in extracting the visual evidence required to inform the category decision is most likely to evoke uncanny experience. The relationship between perceptual difficulty and uncanny experience has yet to be tested.

Greater task difficulty can also be indicated by a longer duration of fixations, that is, dwell time (Buswell, 1935). Dwell time is thought to reflect the demands of actively processing a fixated region of visual interest (for process monitoring models see, Rayner, 1987; Rayner and Fischer, 1987) and of the degree of in-depth processing of task-relevant information (Duncan and Humphreys, 1989; Remington and Folk, 2001; Becker, 2011). For example, longer dwell time can reflect greater processing demands when discriminating between similar stimuli (Shen et al., 2003; Becker, 2011) and when matching observed visual stimulus information to internal representations (Goldberg and Kotval, 1999). Uncertainty in perceptual decision making entails enhanced recruitment of attentional resources for the accumulation of sensory evidence (Heekeren et al., 2008). Longer dwell time might therefore be expected when an individual encounters difficulty extracting perceptual information from a categorically ambiguous stimulus and matching this information to representations of the non-human or human category (Fitts et al., 1950; Barton et al., 2006). Assuming that Mori's hypothesis is correct, the cognitive demands placed on processing perceptual information should be greater at the category boundary of the DHL compared with the other regions of the DHL where there is little or no categorization ambiguity. Based on the preceding considerations, it would be consistent with Mori's hypothesis that differences in perceptual processing demands along the DHL are indicated by differences in dwell time.

The relative importance of particular visual features such as the eyes, nose, and mouth is not addressed in the Uncanny Valley Hypothesis, although these features have received attention in uncanny-related research (e.g., Seyama and Nagayama, 2007; MacDorman et al., 2009a; Looser and Wheatley, 2010). Certain facial features and particularly the eyes do generally draw more overt attention in various tasks (Yarbus, 1967; Fisher and Cox, 1975; Haith et al., 1977; Walker-Smith et al., 1977; Janik et al., 1978; Langdell, 1978; Althoff and Cohen, 1999; Minut et al., 2000;

Henderson et al., 2005; Schwarzer et al., 2005). Evidence of a facial feature hierarchy pertaining to the relative importance of the eyes, nose, and mouth (and of other features) is long established (Ellis et al., 1979; Fraser and Parker, 1986). But a central question affecting both design considerations and the cognitive processing of such realistic humanlike characters is whether the relative importance of these facial features is modulated by categorization ambiguity. In other words, does the relative importance of these facial features change as a function of the difficulty of perceptual decision making at different points along the DHL? Uncanny-related research shows that the relative realism of the eyes can influence the experience of negative affect (Seyama and Nagayama, 2007; MacDorman et al., 2009a), and a recent study suggests that the eyes and mouth contain more diagnostic information for determining whether human and humanlike faces are animate (Looser and Wheatley, 2010). These findings mean that certain facial features might be preferentially fixated in order to determine category membership when sensory evidence is unclear. This would be indicated by a change between categorically unambiguous and ambiguous faces in the relative proportion of fixations to the eyes, nose, and mouth.

An important consideration is that women show a general processing advantage for perceptual details in face detection and facial identity discrimination tasks, especially under more demanding processing conditions (McBain et al., 2009; see also Lewin and Herlitz, 2002; Rehnman and Herlitz, 2007). This advantage is well investigated for face recognition (for a review, see Herlitz and Rehnman, 2008) and facial expression recognition (Kirouac and Doré, 1985; Nowicki and Hartigan, 1988; Thayer and Johnsen, 2000; Hall and Matsumoto, 2004; Montagne et al., 2005; Scholten et al., 2005; Biele and Grabowska, 2006). These studies show that women process faces faster and more accurately than do men. The female advantage in expression recognition is associated with greater female attention to the eyes, indicated by longer eye-related dwell time and larger fixation number (Hall et al., 2010). Similar findings have been found in other tasks and with other stimuli (Miyahira et al., 2000a,b).

The aim of this investigation was to examine whether there are differences in the way visual attention is overtly oriented to the eyes, nose, and mouth of categorically unambiguous and ambiguous faces along the DHL. The DHL was represented using

morph continua (e.g., Hanson, 2006; Seyama and Nagayama, 2007; Ho et al., 2008), the advantage of which is that they allow careful experimental control of differences in humanlike appearance and the exclusion of confounding perceptual dimensions (for a critical overview of the use of morph continua in uncanny research, see Cheetham and Jancke, 2013). The morph continua were generated from avatar and human parent faces and the morphs presented in a forced choice categorization task. Eye measures of fixation number and dwell time were collected to determine, firstly, the presence of a general face feature hierarchy for the eyes, nose, and mouth. Second, we examined whether the relative importance of these features changes between categorically unambiguous avatar and human faces compared with categorically ambiguous faces at the category boundary of the DHL. The categorically unambiguous faces and the peak in categorization ambiguity were determined on the basis of the categorization responses and response times (RT) in the categorization task. Based on the preceding considerations, we predicted that greater categorization ambiguity would be reflected in a shift in the relative importance of facial features in terms of dwell time but not necessarily in terms of the actual location of fixations. Third, we examined whether there are differences between men and women in categorization responses and response latencies and in our measures of eye movement. Based on the available literature, we anticipated that women would generally show shorter response latencies than men and explored the possibility that the eyes are generally more important for women than for men when processing faces along the DHL.

4.3. Experiment 4. Category processing and eye-movement

4.4. Materials and Methods

Participants

All volunteers were healthy students ($N = 60$, 31 females, 29 males; aged 18–32, mean 23 years) of the University of Zurich, with no record of neurological or psychiatric illness and no current medication. All participants had normal or corrected-to-normal visual acuity, were native or fluent speakers of Swiss or Standard German, and consistently right-handed (Annett, 1970). All participants

reported having no active design or gaming experience with computer-generated characters in video games, virtual role-playing games, second life and other virtual reality environments. One female participant was omitted from analyses because she did not show a logistic component in the response function of her data (see section Logistic Function of Categorization Responses). Heavy track loss, that is, failure by the eye-tracking system to detect the pupil (mostly in this study because of excessive blinking) resulted in there being no fixation data in over 20% of trials in another female and in five male participants. The data of these participants was excluded from further analysis. Written informed consent was obtained before participation according to the guidelines of the Declaration of Helsinki. Each volunteer received 15 Swiss Francs for participation. The study and all procedures and consent forms were approved by the Ethics Committee of the University of Zurich.

Stimuli

Ten morph continua were generated using Morpher 3.3 software (Zealsoft, Inc., Eden Prairie, MN, USA) to represent the DHL. The parent images of the continua were natural human and avatar faces. Each face was male, indistinctive, presented with full frontal view, direct gaze, neutral expression and no other salient features (e.g., facial hair, jewelry). The avatars were generated using the avatar modeling software Poser 7 (Smith Micro Software, <http://www.smithmicro.com>). The avatars' facial geometry and texture (configural cues, skin tone) were modeled to closely match the human counterpart. The parent images were then edited using Adobe Photoshop CS3. The external features of each parent image were masked with a generic elliptic form and black background, and contrast levels, overall brightness of the parent images of each continuum were adjusted to match. Each morph continuum comprised 15 different morphed images, each morph separated by an increment of 6.66% in physical difference. A pilot study of $N = 20$ participants who were not recruited for the present study was conducted using the morph images of each morph continua in a forced choice categorization task. These images were numbered 1 (avatar end) to 15 (human end). The purpose of the pilot study was to verify that the morph position along the continua of the category boundary and the shape of the response function of each continuum was consistent across all continua before performing the present eye-tracking study. This was confirmed. The task and

analyses for the pilot study were the same as described in the following sections for the present study. Based on the pilot study, the morph images 3–15 (i.e., the 13 most human-like morphs) of each morph continuum were used in the present study and re-labeled 1 (avatar end) to 13 (human end; Figure 3). This was done to achieve a better balance in the number of morph images either side the category boundary. At a viewing distance of 62 cm, the stimuli (400 × 500 pixels) subtended a visual angle of 11° × 14°. This is approximately equivalent to viewing a real face from a normal distance during conversation of 90–100 cm (Hall, 1966; Henderson et al., 2005).

Design and Procedure

The participants were examined individually, each examination lasting approximately 20 min in total. The participant received general instructions before beginning each experimental phase (i.e., calibration of eye-tracker, practice pre-test, and main test). The participant was seated in a chair with a chin rest and head band to restrain head movement. An I-View SMI dark pupil remote eye-tracker system (SensoMotoric Instruments, GmbH) with sampling frequency at 50 Hz and up to 0.5° accuracy was used to record horizontal and vertical eye positions. A 13-point calibration procedure was applied at the beginning of the experiment and drift correction was performed automatically. After the calibration procedure, the participant read written instructions on the monitor before commencement of the forced choice categorization task. This began with a practice pre-test of five trials using stimuli drawn from an unused continuum. Comprehension of the task and correct use of the response buttons was ensured. The participant then pressed the response box to initiate trial presentation for the main test.

Each trial started with a blank screen for 2200 ms to allow rest and blinks followed by a fixation point for 800 ms (prepare signal) and then a stimulus. Participants were instructed to fixate on the point until the stimulus appeared, to view the stimulus freely and to identify the stimulus quickly and accurately as either an avatar or human by pressing one of two response keys. The stimulus disappeared from view following button press or after the maximum viewing duration of 3000 ms. Participants performed 130 trials, with each morph of each continuum being presented once, individually, in the centre of the monitor and in random order.

The task was conducted in a sound attenuated and light-dimmed room, and stimuli were presented on a LCD monitor (1280 × 1024 resolution, 60 Hz refresh rate, at eye-to-monitor viewing distance of 62 cm), using Presentation® software (Version 14.1, www.neurobs.com).

Analyses

Eye movement data from stimulus onset until participant response was analyzed. BeGaze software Version 2.5 (SensoMotoric Instruments, GmbH) was used to pre-process the data. A fixation was defined as consecutive eye gaze positions within an area of 1° for a period of 100 ms or more. Blinks and any fixations that fell outside the masked face region were discarded from data analysis. Four areas of interest (AOI) were assigned: the internal face area as a whole and the individually defined areas of the eyes (including the eyebrows), nose, and mouth, as shown in Figure 2. Total fixation number and cumulative total fixation duration for each AOI was computed. The same AOI were applied to all stimuli to ensure consistency in the definition of boundaries between facial features and size of AOI. Given the different response latencies, values were normalized to represent the “proportion of the total number of fixations” (referred to in the following as “fixations”) and the “proportion of the total fixation duration” (referred to in the following as “dwell time”) within each AOI and entered into statistical analyses as dependent variables. For all analyses, only data was included for trials in which responses were made and before stimulus duration elapsed. All data analyses were performed using SPSS version 17.0 (SPSS, Inc., Chicago, IL, USA).

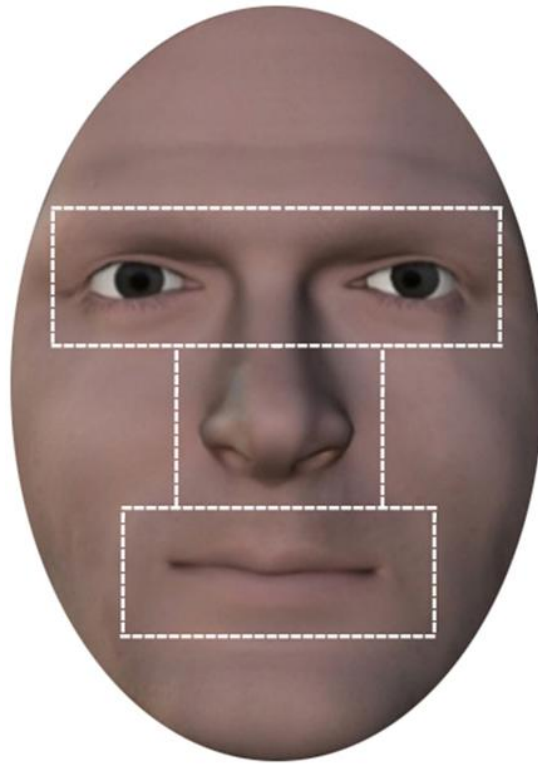


Figure 2. Areas of interest. Illustration of internal region of the face comprising the areas of interest for the facial features eyes, nose, mouth.

4.5. Results

Forced Choice Categorization

Forced choice – categorization responses

The slope of the categorization response function was used to summarize the avatar-human categorization response data. The function was determined by fitting logistic function models to the data of each participant across continua. Informal inspection of each participant's fitted regression curve indicated for each participant a response function with a sigmoid shape indicative of a category boundary, with the exception of one participant who following the analysis of logistic function (see section Logistic Function of Categorization Responses) was excluded from subsequent analyses. Parameter estimates were derived from the model and entered in the analyses of the logistic function of categorization responses and of the category boundary.

Figure 3 shows the mean aggregated response data for the 10 continua across male and female participants and, for purposes of illustration, the fitted response function computed on the basis of the grand mean of these continua. The sigmoid-shaped curve of the fitted response function shows a lower and upper asymptote of avatar and human categorization responses that nears 100% for avatars and 95% for humans, respectively.

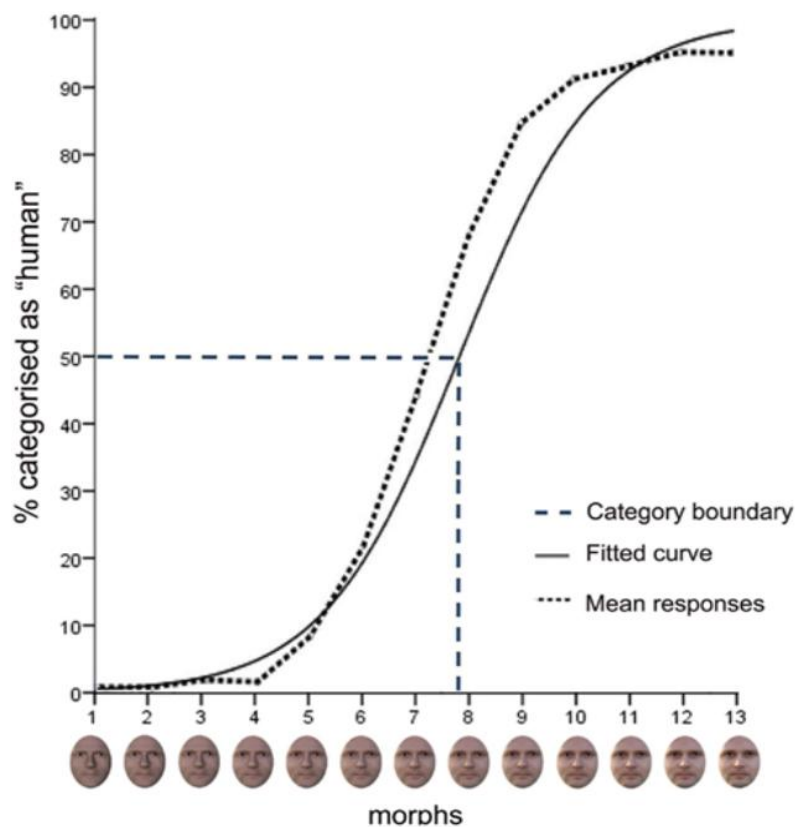


Figure 3. Mean responses in the categorization task. The mean aggregated response (dotted line), shown in terms of the percentage of “human” responses, and the mean logistic curve (continuous line) across all continua and participants are displayed. The mean category boundary (dashed line) indicates the morph position along the continua at which there is maximum uncertainty of 50% in categorization judgments. The bottom panel of the figure shows an example of a morph continuum.

Logistic function of categorization responses

The derived parameter estimates for the logistic function of data averaged across the 10 continua for each participant were tested against zero in a one-sample *t*-test. The

result showed a highly significant logistic component, $t_{58} = 20.19$, $p > 0.001$, that captures a sigmoid-shaped function consistent with a category boundary (Harnad, 1987; Figure 3).

Category boundary

The estimates for the β_0 and β_1 parameters of each participant across continua were used to compute the *category boundary value* [i.e., $y = 0.5$: $-\ln(\beta_0)/\ln(\beta_1)$]. This value indicates the morph position along the continua that corresponds with the ordinate midpoint between the lower and upper asymptotes, that is, the point of maximum uncertainty of 50% in categorization judgments. Across continua, the mean category boundary value ($M = 7.81$, $SD = 0.83$) corresponds with face morph position 8 (Figure 3). An independent t -test showed that the mean category boundary value for male participants ($M = 8.23$, $SD = 0.93$) was not significantly different from that for female participants ($M = 7.49$, $SD = 0.68$).

Categorization response as a function of morph position

We tested for expected differences in the categorization responses as a function of morph position (i.e., 1–13) and examined potential differences in categorization responses between the continua and male and female participants. A repeated measures of analysis of variance (RM-ANOVA) was performed on the dependent variable “response” (for avatar-human categorizations) of each participant with the “morph position” (13 levels) across the 10 continua. “Sex” was entered in analysis as between subject variable. Greenhouse–Geisser adjustment was applied to correct the degrees of freedom for violation of the sphericity assumption (and applied as appropriate in subsequent analyses). The analysis showed a highly significant effect for morph position [$F(4.39, 259.23) = 765.48$, $p < 0.001$].

Forced choice – categorization response times

Differences in category decision difficulty, as indicated by the slope of the categorization response function, are likely to be reflected in different response latencies for the morphs of the continua. To gain an overall picture of the RT at

different morph positions, a one-way RM-ANOVA with *morph position* (13 levels) and RT as the dependent variable was conducted. “Sex” was entered in analysis as between subject variable. The analysis showed a highly significant effect for morph position [$F(4.28, 252.3) = 168.2, p < 0.001$], and for sex, $F(4.28, 252.3) = 4.54, p = 0.001$ (Figure 4).

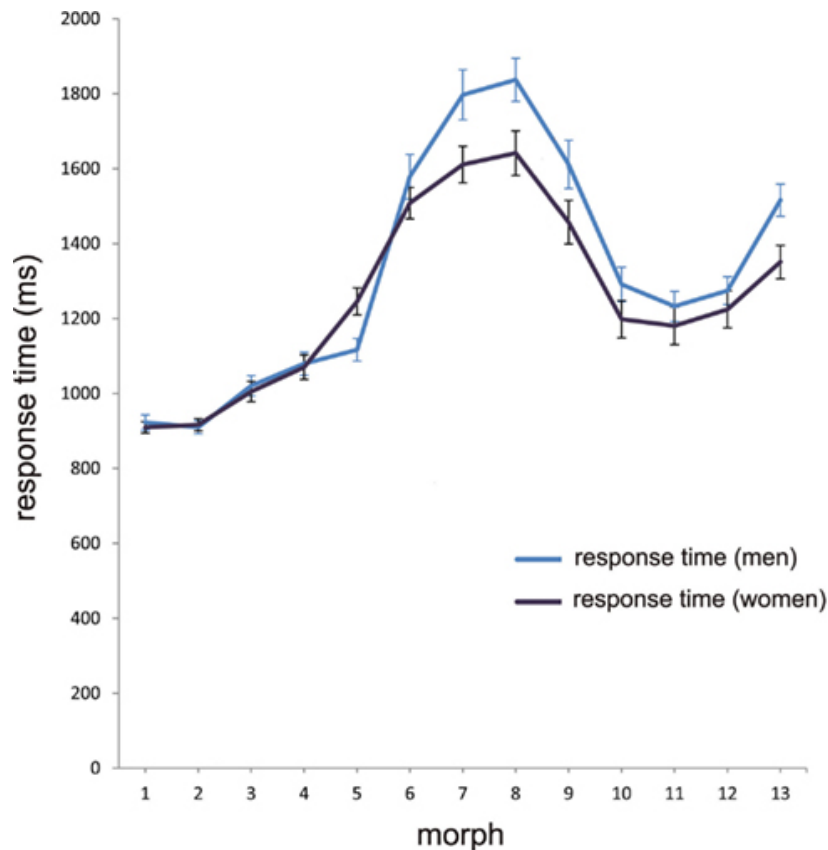


Figure 4. Mean RT in the categorization task. The mean aggregated response time for men (blue line) and women (black line) at the different morph positions 1–13, with error bars as standard errors ($N = 59$). Note that the longest response latencies for men and women at morph 8 correspond with the category boundary shown in Figure 2.

The longest response latency should correspond with the position of the category boundary. For the whole sample, the category boundary is morph 8. To characterize the effect more clearly, the mean RT values at morph 8 were compared with the mean RT values at all other morph positions. A one-way RM-ANOVA analysis with *morph position morph 8 versus all other morphs* and RT as dependent variable and sex as between subject variable showed that RT at morph 8 ($M = 1.73, SD = 0.33$) differed highly significantly from RT for the other morph positions ($M = 1.25, SD = 0.16$), $F(1, 58) = 158.93, p < 0.001$. There was no effect for sex.

Given the interest in asymmetry in performance for within-category avatar and human faces, we compared mean RT for morph positions 1–4 with mean RT for the morph positions 10–13. These morph positions were unambiguously assigned to one or other category with a decision certainty of in excess of 90%; a much less conservative criterion is often used in CP research (e.g., 66% as in Ectoff and Magee, 1992; Beale and Keil, 1995). For this, a one-way RM-ANOVA was conducted using the factor *morph position* (two levels: 1–4, 10–13) and RT as dependent variable. This indicated that mean RT for avatars at morph positions 1–4 ($M = 1.02$, $SD = 0.12$) was significantly shorter than that for the human faces at morph positions 10–13 ($M = 1.33$, $SD = 0.24$), $F(1, 58) = 135.29$, $p < 0.001$. There was also a gender effect, such that males were slower to humans faces ($M = 1.38$, $SD = 0.21$) than females ($M = 1.28$, $SD = 0.26$), $F(1, 58) = 5.14$, $p = 0.027$ (Figure 4). There was no difference in RT between man and women for faces of the avatar category.

Eye Movement

Internal face area, and dwell time

To determine the relative importance of the internal face area comprising the eyes, nose, and mouth as a whole, a one-way RM-ANOVA was conducted using the factor *morph position* (13 levels: 1–13) and dwell as dependent variable. Sex was entered in the analysis as between subject variable. No significant differences were found.

Eyes, nose, mouth, and dwell time

To examine the relative role of eyes, nose, and mouth for the *DHL regions* avatar category, human category and for the morph of peak uncertainty (i.e., morph position 8) during category decision making, a two-way RM-ANOVA was conducted with the factors *DHL region* (3 levels: 1–4, 8, 10–13) and feature (three levels: eyes, nose, mouth) and dwell as dependent variable. “Sex” was entered in analysis as between subject variable. The analysis showed a general difference between men and women in the relative importance of the eyes, nose, and mouth, $F(68.1, 1.36) = 4.47$, $p < 0.027$. The pre-planned contrasts show that there were generally no significant

differences for the mouth between men and women, while the eyes were more salient relative to the nose for women ($M_{\text{eyes}} = 58.55$, $SE_{\text{eyes}} = 3.39$; $M_{\text{nose}} = 29.55$, $SE_{\text{nose}} = 2.95$) than for men ($M_{\text{eyes}} = 48.78$, $SE_{\text{eyes}} = 4.36$; $M_{\text{nose}} = 33.38$, $SE_{\text{nose}} = 3.83$), $F(1, 50) = 5.26$, $p < 0.026$ (Figure 5).

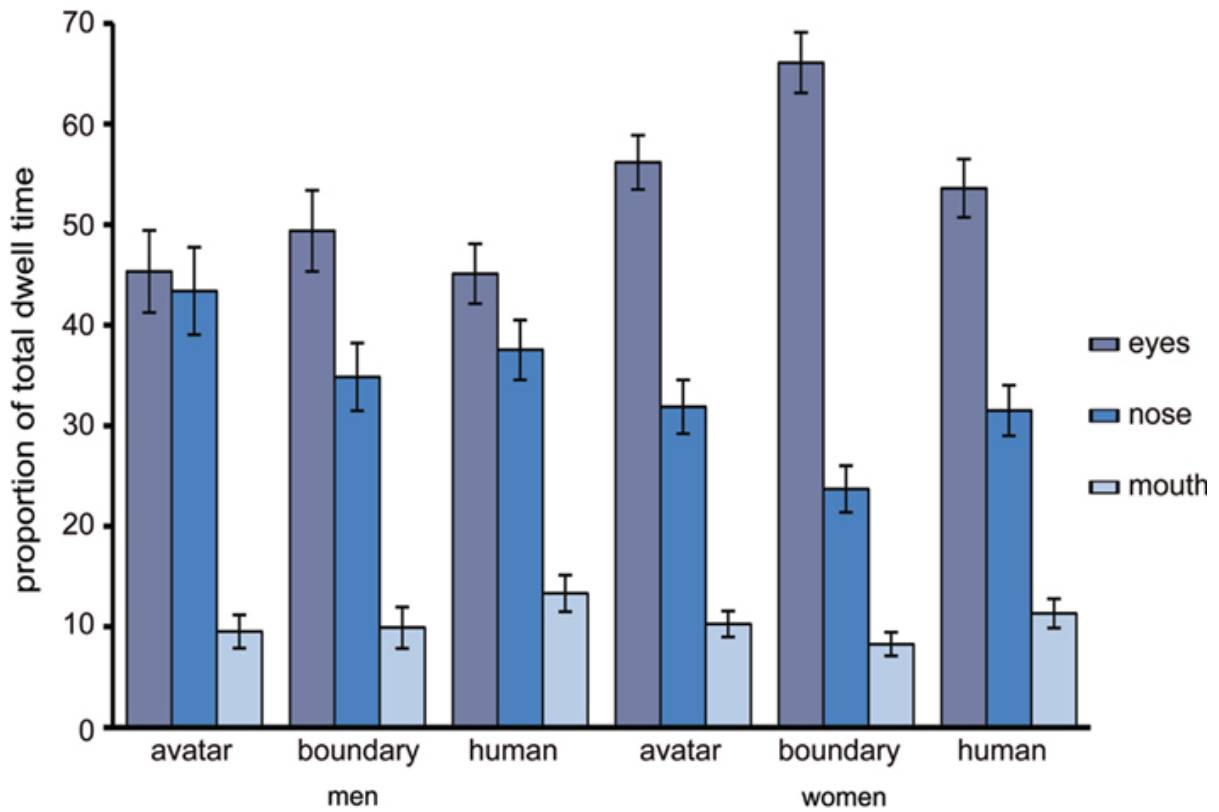


Figure 5. The proportion of total fixation duration on the eyes, nose, and mouth. The relative proportion of total dwell time to the eyes, nose, and mouth for the avatar and human categories and for the peak in uncertainty at the category boundary, with error bars as standard errors ($N = 53$).

The analysis showed also a significant interaction effect for ENM \times DHL region [$F(2.65, 132.46) = 4.24$, $p = 0.009$] such that the relative salience of the eyes, nose, and mouth varies as a function of DHL region. At the peak compared with the avatar category, pre-planned contrasts show a significant increase in sampling of the eyes ($M_{\text{peaks}} = 57.56$, $SE = 3.66$; $M_{\text{avatar}} = 50.34$, $SE = 3.42$) compared with the nose ($M_{\text{peak}} = 29.38$, $SE = 2.31$; $M_{\text{avatar}} = 37.84$, $SE = 3.16$), $F(1, 51) = 7.78$, $p < 0.007$, and a significant decrease in sampling of nose ($M_{\text{peak}} = 29.38$, $SE = 2.31$; $M_{\text{avatar}} = 37.84$, $SE = 3.16$) compared with the mouth ($M_{\text{peak}} = 8.43$, $SE = 1.17$; $M_{\text{avatar}} = 10.62$, $SE =$

1.45), $F(1, 51) = 0.78, p < 0.033$. At the human compared with the avatar category, there was a significant decrease in sampling of the nose ($M_{\text{human}} = 34.65, SE = 2.74$; $M_{\text{avatar}} = 37.84, SE = 3.62$) relative to the mouth ($M_{\text{human}} = 11.62, SE = 1.43$; $M_{\text{avatar}} = 10.07, SE = 1.45$), $F(1, 51) = 6.96, p < 0.011$, but no significant effect for eyes.

Internal face area, and fixation number

A one-way RM-ANOVA was conducted using the factor *morph position* (13 levels: 1–13) and dwell as dependent variable, and with sex as between subject variable. There were no effects.

Eyes, nose, mouth, and fixation number

To examine the relative role of eyes, nose, and mouth for the three DHL regions, as in the preceding analysis, a two-way RM-ANOVA was conducted with the factors *DHL region* (three levels: 1–4, 8, 10–13) and feature (three levels: eyes, nose, mouth) and fixations as dependent variable. “Sex” was entered in analysis as between subject variable. The analysis showed a highly similar general pattern for fixation as for dwell. There was a main effect for facial feature such that the eyes, nose, and mouth are differently important, $F(1.53, 76.53) = 58.91, p < 0.001$. The pre-planned contrasts show that the eyes were more salient relative to the nose ($M_{\text{eyes}} = 54.01, SE_{\text{eyes}} = 3.23$; $M_{\text{nose}} = 31.67, SE_{\text{nose}} = 2.71$) [$F(1, 51) = 116.61, p < 0.001$] and mouth ($M_{\text{mouth}} = 11.55, SE_{\text{mouth}} = 1.93$) $F(1, 51) = 21.63, p < 0.001$ and the nose more salient than the mouth, $F(1, 51) = 53.12, p < 0.001$ (Figure 6). But there was no significant effect for DHL region, though this did approach significance, $F(148.9, 74.44) = 3.27, p = 0.058$. There were no other significant effects for fixations.

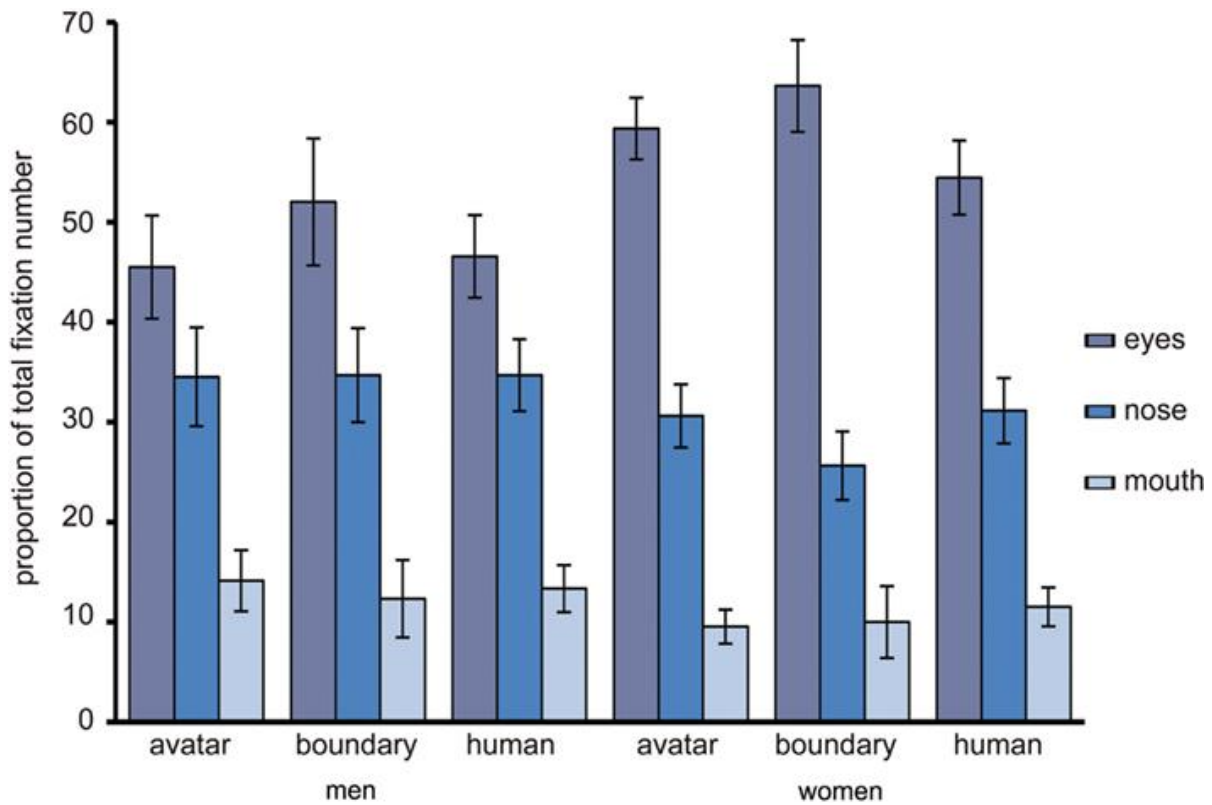


Figure 6. The proportion of total fixation number to the eyes, nose, and mouth. The relative proportion of total fixation number to the eyes, nose, and mouth for the avatar and human categories and for the peak in uncertainty at the category boundary, with error bars as standard errors ($N = 53$).

4.6. Discussion

The unique feature of uncanny research is the interest in understanding the impact on subjective experience and behaviour of uncertainty in distinguishing highly realistic non-human objects (or specific perceptual attributes of these) from their human counterpart (Ramey, 2005; MacDorman et al., 2009b). In terms of categorization behavior, ambiguity in assigning category membership should be greatest at the category boundary (e.g., Campbell et al., 1997). For our avatar-human continua, a category boundary was indicated by the sigmoid shape of the response function (Harnad, 1987). This shape reflects a monotonic increase in categorization accuracy (i.e., the proportion of decision responses in favour of the avatar or human category) and a corresponding monotonic decrease in response latency as a function of morph distance from the category boundary. The data thus indicate that the category boundary marks the point of greatest cognitive conflict between the two competing categorization response tendencies.

Overt orientation of visual attention to the eyes, nose, and mouth was therefore compared between faces at the category boundary and faces that were unambiguously assigned to the avatar or human category. The data for the number of fixations showed no significant differences in this comparison, indicating that category ambiguity has no specific effect on the targeting of fixations to these facial features. But this data did confirm that the relative importance of features in the face feature hierarchy (e.g., Walker-Smith et al., 1977; Althoff and Cohen, 1999) is similar for faces of variously humanlike appearance along the DHL, the eyes being especially important. The latter might be expected given the high human sensitivity to eye information (Itier et al., 2007; Itier and Batty, 2009) and the diagnostic value of the eye region for various face processing tasks (e.g., Althoff and Cohen, 1999; Hall et al., 2010). This eye dominance might also reflect the influence of an automatic gaze strategy that entails positioning fixations near a center of gravity (Bindemann et al., 2009). This center of gravity is biased toward a position between the eyes and nose (e.g., Deaner and Platt, 2003; Grosbras et al., 2005; Tyler and Chen, 2006; Hsiao and Cottrell, 2008; Saether et al., 2009; van Belle et al., 2010). It is possible that the presentation of full frontal faces in this study facilitated the targeting of the eyes (Saether et al., 2009). On the other hand, fixations to the eye region might have been advantageous in our categorization task because this point of regard might best serve rapid visual inspection of the faces, with the eyes being processed foveally and the nose and other areas parafoveally. Compared with our full frontal face stimuli, a change in the relative viewpoint of the observer might alter the fixation patterns generally. This is likely because head (and body) orientation can influence the direction of the viewer's attention (Hietanen, 1999, 2002; Langton, 2000; Pomianowska et al., 2011). The impact of different viewpoints on eye movement patterns when processing categorically unambiguous and ambiguous faces along the DHL is not known.

The data for dwell time revealed a different picture. Categorization ambiguity influenced the relative amount of dwell time spent extracting visual information from the eye, nose, and mouth before making a category decision. This effect was associated with the non-human side of the DHL, with a relative shift away from the nose region in avatars to the regions of the eyes and mouth of faces at the category

boundary. A similar effect almost reached significance for the comparison of the human versus ambiguous faces. The eye movement data thus suggest that avatars are processed differently than are ambiguous faces. The behavioural data support this by showing that category assignment of avatar faces is also much faster in comparison with the assignment of human faces. This replicates Cheetham et al.'s (2011) finding, using a much larger participant sample in the present study. Given that our participants' perceptual and categorization experience with novel non-human but highly humanlike faces generally and especially with those presented in this study does not compare to their everyday expertise with human faces (Diamond and Carey, 1977; Tanaka and Curran, 2001; see, Ramey, 2005), one might have expected that the assignment of faces to the human category would have been easier than to the avatar category.

One possible explanation for the strong difference in RT for the avatar compared with the human category decisions is that this decision is influenced by a strategy that involves the detection of perceptual information that is diagnostic of the non-human category, that is, of the category of which the human participant is not a member (see the race-feature hypothesis, Levin, 2000). This would mean that classification decisions are based on establishing the presence or absence of avatar-specifying perceptual information, with faces being coded and categorized in terms of "avatar or not avatar" rather than as "avatar or human." Assuming that detecting the presence rather than the absence of perceptual information is cognitively less demanding, this classification strategy would result in a classification advantage in RT for avatar faces. This avatar-feature hypothesis thus suggests that there might be preferential processing of avatar-specifying information during explicit categorization of categorically unambiguous avatars.

The question is what perceptual information might be diagnostic for avatar faces during categorization. One possibility, given the task's context of processing novel avatar faces and faces of the familiar human category, is that perceptual information indicating novelty could in itself serve as a readily identifiable primitive feature of members of the avatar category (cf. Levin, 2000). Certain properties of an avatar's face such as the general shading of the smoothed skin texture might be relatively easy to detect and support a fast and reliable strategy for categorization (cf.

Cheetham et al., 2011). Alternatively, avatar-specifying information might relate to more detailed shape and texture information of the eyes, nose, and mouth (MacDorman et al., 2009a; Looser and Wheatley, 2010) independently of novelty processing. But it should be noted that visual discrimination performance of facial texture and shape properties is influenced by experience (Vuong et al., 2005; Balas and Nelson, 2010) and that perceptual and categorization experience can influence the selection of perceptual details used for analysis before categorization (e.g., Schyns, 1991; Schyns and Murphy, 1994). Given that novel stimuli are known to evoke different patterns of fixations than familiar stimuli (Althoff and Cohen, 1999), it is possible that repeated exposure to and greater perceptual and categorization experience with non-human faces might result in a change in eye movement patterns. Participants were therefore selected with a view to limiting the potential impact on eye movement behaviour of active experience with building and modifying avatars in video games and virtual reality environments and of active gaming experience, while recognizing that active and incidental exposure to humanlike characters in other media (e.g., comics and books, movies, and commercials) could influence eye movement patterns. The influence on eye movement patterns of active and incidental perceptual and categorization learning with humanlike characters along the DHL awaits investigation.

Our main finding of the present study concerns the relative shift away from the nose region in avatars to the regions of the eyes and mouth of ambiguous faces. One approach to understanding which perceptual details are perceived and analyzed before categorization of these faces might be to consider the role of spatial frequency information. Eye movements are selective for spatial frequency information (e.g., Tavassoli et al., 2009) and the nose region is more diagnostic for processing stimulus information in coarser spatial frequency scales (Schyns et al., 2002). Coarser frequencies (i.e., low spatial frequencies) are suggested to facilitate encoding of shape and texture on the basis of the general luminance properties and shading of the face, corresponding therefore to a faster and broader but less detailed processing of the face (Sergent, 1986; Schyns and Oliva, 1994; Schyns and Gosselin, 2003). Given the relative shift in dwell time toward the nose region of avatars and the comparatively short RT for avatars, it is conceivable that a broad processing strategy

would help the quick identification of perceptual information that indicates membership of the avatar category.

In contrast, longer response latencies and a greater proportion of dwell time to the eyes and mouth of ambiguous faces might be interpreted as reflecting a shift away from a faster but broad and less detailed processing strategy to a more time-consuming strategy of processing finer perceptual details (Schyns and Murphy, 1994; Lamberts, 1998; Johansen and Palmeri, 2002). This is especially likely for highly similar faces (e.g., Oliva and Schyns, 1997) and, presumably, for faces that are difficult to discriminate in terms of avatar or human category membership. The eye region is more diagnostic in finer frequency scales (Schyns et al., 2002) and the processing of the specific shape of the mouth and eyes and of the contours of the nose benefits from a finer spatial resolution (i.e., high spatial frequencies; Sergent, 1986; Schyns and Gosselin, 2003). We did not examine or manipulate spatial frequency, but these considerations suggest that there may have been a subtle shift in the relative salience of the nose of avatars to the eyes and mouth of ambiguous faces in a manner that resembles a course-to-fine sampling strategy along the DHL. This strategy would entail rapid and coarser perceptual information processing that is sufficient for avatar categorization and more time-consuming processing of finer details (in addition to the coarser details) in order to disambiguate the category membership of faces near and at the category boundary. This course-to-fine hypothesis might be of further interest for uncanny-related research.

Women were faster than men to make category decisions. This is generally consistent with the reported female advantage in other face processing categorization tasks (e.g., Hall et al., 2010). But this finding applied only to ambiguous and human faces. In fact, the greatest difference between men and women in RT was at the category boundary. The category boundary is associated with enhanced perceptual discrimination ability along the DHL (Cheetham et al., 2011; Cheetham and Jancke, 2013), but these studies did not examine gender-related differences. Like the present study, these studies presented male faces only. Whether the findings of this study similarly apply for female stimuli awaits further investigation. Women do show a processing advantage for discriminating perceptual details in faces, especially under more demanding processing conditions (McBain et

al., 2009; see also McGivern et al., 1998; Kimchi et al., 2008), but there were no gender-related differences in fixation or dwell time between avatar, human, and ambiguous faces, suggesting that any female advantage in RT does not translate into differences in eye movements along the DHL. The only gender-related difference in eye movements was in the general face feature hierarchy, with enhanced female attention to the eye region across the DHL for dwell time. A similar finding in women has been shown for expression recognition, and this was associated with greater eye-related dwell time and fixation number (Hall et al., 2010). Enhanced female attention to the eyes could be considered a task-independent effect associated with greater processing of socially relevant information in women (Saether et al., 2009).

In summary, these findings show that categorization ambiguity influenced the relative amount of dwell time spent extracting visual information from the eyes, nose, and mouth, though in comparison only with the non-human side of the DHL. Together with the shorter decision response latencies for avatar faces, the findings thus suggest that categorically unambiguous avatars are processed differently than ambiguous or human faces. Guided in part by current uncanny research, the focus of this first eye movement study of category decision making along the DHL was placed on the eyes, nose, and mouth. These internal face features were also selected because the greater part of eye sampling behavior was directed toward these regions in previous studies (e.g., Barton et al., 2006). But the arbitrary size and shape of the AOI used in the present study might well have masked potential findings. For example, our data show a consistent 70–75% of fixations and dwell time within the internal facial regions of all faces along the DHL, suggesting that the external face area is nevertheless relevant for avatar-human category decision making. External facial features such as the jaw line, hair profile, and head contour, which were masked in this study, might be informative in further studies of the DHL and category processing. We focused on highly realistic non-human faces, and our findings might apply for similarly highly realistic androids (see Saygin et al., 2012) but not necessarily for less humanlike avatars, cartoon faces, and robots (see Chen et al., 2010; Saygin et al., 2012). Our choice of stimuli did not reflect the examples of humanlike objects in the depiction of Mori's hypothesis in Figure 1. This was because we sought to ensure careful experimental control and exclusion of visual cues of other perceptual dimensions that could confound with judgments of human likeness

along the DHL (Cheetham and Jancke, 2013). We used 2D facial stimuli, but given the growing impact of 3D technology in various media, a promising avenue of further research would be to examine the effect of stereoscopic depth cues in 3D compared with 2D facial stimuli on eye sampling behavior and categorization (e.g., Bühlhoff and Newell, 2000; Liu and Ward, 2006; Chelnokova and Laeng, 2011).

Conflict of Interest Statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgments

This work was supported by the European Union FET Integrated Project PRESENCCIA (Contract number 27731).

4.7. References

- Althoff, R. R., and Cohen, N. J. (1999). Eye-movement-based memory effect: a reprocessing effect in face perception. *J. Exp. Psychol. Learn. Mem. Cogn.* 25, 997–1010.
- Annett, M. (1970). A classification of hand preference by association analysis. *Br. J. Psychol.* 61, 303–321.
- Balas, B., and Nelson, C. A. (2010). The role of face shape and pigmentation in other-race face perception: an electrophysiological study. *Neuropsychologia* 48, 498–506.
- Bartneck, C., Kanda, T., Ishiguro, H., and Hagita, N. (2007). “Is the Uncanny Valley an Uncanny Cliff?” in *Proceedings of the 16th IEEE, RO-MAN*, Jeju, 368–373.
- Barton, J. J. S., Radcliffe, N., Cherkasova, M. V., Edelman, J., and Intriligator, J. M. (2006). Information processing during face recognition: the effects of familiarity, inversion, and morphing on scanning fixations. *Perception* 35, 1089–1105.
- Beale, J. M., and Keil, F. C. (1995). Categorical effects in the perception of faces. *Cognition* 57, 217–219.
- Becker, S. I. (2011). Determinants of dwell time in visual search: similarity or perceptual difficulty? *PLoS ONE* 6:e17740.
- Biele, C., and Grabowska, A. (2006). Sex differences in perception of emotion intensity in dynamic and static facial expressions. *Exp. Brain Res.* 171, 1–6.
- Bindemann, M., Scheepers, C., and Burton, A. M. (2009). Viewpoint and center of gravity affect eye movements to human faces. *J. Vis.* 9, 1–16.
- Bühlhoff, I., and Newell, F. N. (2000). Investigating categorical perception of gender with 3-D morphs of familiar faces. *Perception* 29, 57.

- Buswell, G. T. (1935). *How People Look at Pictures*. Chicago: University of Chicago Press.
- Campbell, R., Pascalis, O., Coleman, M., Wallace, S. B., and Benson, P. J. (1997). Are faces of different species perceived categorically by human observers? *Proc. R. Soc. Lond. B Biol. Sci.* 264, 1429–1434.
- Cheetham, M., Suter, P., and Jancke, L. (2011). The human likeness dimension of the “uncanny valley hypothesis”: behavioral and functional MRI findings. *Front. Hum. Neurosci.* 5:126.
- Cheetham, M., and Jancke, L. (2013). Perceptual and category processing of the Uncanny Valley Hypothesis’ dimension of human likeness: some methodological issues. *J. Vis. Exp.* e4375..
- Chelnokova, O., and Laeng, B. (2011). Three-dimensional information in face recognition: an eye-tracking study. *J. Vis.* 11, 27.
- Chen, H., Russell, R., Nakayama, K., and Livingstone, M. (2010). Crossing the “uncanny valley”: adaptation to cartoon faces can influence perception of human faces. *Perception* 39, 378–386.
- Cook, M. (1978). “Eye movements during recognition of faces,” in *Practical Aspects of Memory*, eds M. M. Gruneberg, P. E. Morris, and R. N. Sykes (New York: Academic Press). 286–292.
- Deaner, R. O., and Platt, M. L. (2003). Reflexive social attention in monkeys and humans. *Curr. Biol.* 13, 1609–1613.
- Desimone, R., and Duncan, J. (1995). “Neural mechanisms of selective visual attention.” *Annu. Rev. Neurosci.* 18, 193–222.
- Diamond, R., and Carey, S. (1977). Development changes in the representation of faces. *J. Exp. Child. Psychol.* 23, 1–22.
- Dill, V., Flach, L., Hocevar, R., Lykawka, C., Musse, S., and Pinho, M. (2012). “Evaluation of the uncanny valley in CG characters,” in *Intelligent Virtual Agents Lecture Notes in Computer Science*, eds N. Yukiko, M. Neff, A. Paiva, and M. Walker (Berlin: Springer), 511–513.
- Duncan, J., and Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychol. Rev.* 96, 433–458.
- Ectoff, N. L., and Magee, J. J. (1992). Categorical perception of facial expressions. *Cognition* 44, 227–240.
- Egeth, H. E., and Yantis, S. (1997). Visual attention: control, representation, and time course. *Annu. Rev. Psychol.* 48, 269–297.
- Ellis, H. D., Shepherd, J. W., and Davis, G. M. (1979). Identification of familiar and unfamiliar faces from internal and external features: some implications for theories of face recognition. *Perception* 8, 431–439.
- Fabri, M., Moore, D. J., and Hobbs, D. J. (2004). “Mediating the expression of emotion in educational collaborative virtual environments: an experimental study,” in *International Journal of Virtual Reality*, London: Springer Verlag.
- Fisher, G. H., and Cox, R. L. (1975). Recognizing human faces. *Appl. Ergon.* 6, 104–109.
- Fitts, P. M., Jones, R. E., and Milton, J. L. (1950). Eye movements of aircraft pilots during instrument-landing approaches. *Aeronaut. Eng. Rev.* 9, 24–29.
- Fraser, I. H., and Parker, D. M. (1986). “Reaction time measures of feature saliency in a perceptual integration task,” in *Aspects of face processing*, eds H. D. Ellis, M. A. Jeeves, F. Newcombe, and A. Young (Dordrecht: Martinus Nijhoff), 45–58.
- Goldberg, J. H., and Kotval, X. P. (1999). Computer interface evaluation using eye movements: methods and constructs. *Int. J. Ind. Ergon.* 24, 631–645.

- Goldberg, J. H., and Wichansky, A. M. (2003). "Eye tracking in usability evaluation: a practitioner's guide," in *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research*, eds J. Hyönä, R. Radach, and H. Deubel (Amsterdam: Elsevier), 493–516.
- Grosbras, M.-H., Laird, A., and Paus, T. (2005). Human cortical regions involved in gaze production, attention shifts and gaze perception: a brain imaging meta-analysis. *Hum. Brain Mapp.* 25, 140–154.
- Haith, M. M., Bergman, T., and Moore, M. J. (1977). Eye contact and face scanning in early infancy. *Science* 198, 853–855.
- Hall, E. T. (1966). *The Hidden Dimension*. Garden City, NY: Doubleday.
- Hall, J. A., and Matsumoto, D. (2004). Gender differences in judgments of multiple emotions from facial expressions. *Emotion* 4, 201–206.
- Hall, J. K., Hutton, S. B., and Morgan, M. J. (2010). Sex differences in scanning faces: does attention to the eyes explain female superiority in facial expression recognition? *Cogn. Emot.* 24, 629–637.
- Hanson, D. (2006). "Exploring the aesthetic range for humanoid robots", in *Proceedings of the ICCS/CogSci-2006 Long Symposium: Toward Social Mechanisms of Android Science*, Vancouver, 39–42.
- Harnad, S. R. (1987). "Introduction: psychological and cognitive aspects of categorical perception: a critical overview," in *Categorical Perception: The Groundwork of Cognition*, ed. S. R. Harnad (New York: Cambridge University Press), 1–25.
- Heekeren, H. R., Marrett, S., and Ungerleider, L. G. (2008). The neural systems that mediate human perceptual decision making. *Nat. Rev. Neurosci.* 9, 467–479.
- Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends Cogn. Sci. (Regul. Ed.)* 7, 498–504.
- Henderson, J. M. (2007). Regarding scenes. *Curr. Dir. Psychol. Sci.* 16, 219–222.
- Henderson, J. M., Williams, C. C., and Falk, R. J. (2005). Eye movements are functional during face learning. *Mem. Cognit.* 33, 98–106.
- Herlitz, A., and Rehnman, J. (2008). Sex differences in episodic memory. *Curr. Dir. Psychol. Sci.* 17, 52–56.
- Hietanen, J. K. (1999). Does your gaze direction and head orientation shift my visual attention? *Neuroreport* 10, 3443–3447.
- Hietanen, J. K. (2002). Social attention orienting integrates visual information from head and body orientation. *Psychol. Res.* 66, 174–179.
- Ho, C.-C., and MacDorman, K. F. (2010). Revisiting the uncanny valley theory: developing and validating an alternative to the Godspeed indices. *Comput. Hum. Behav.* 26, 1508–1151.
- Ho, C.-C., MacDorman, K. F., and Pramono, Z. A. D. (2008). "Human emotion and the uncanny valley: a GLM, MDS, and isomap analysis of robot video ratings," in *ACM/IEEE International Conference on Human–Robot Interaction*, Amsterdam, 169–176.
- Hsiao, J. H., and Cottrell, G. (2008). Two fixations suffice in face recognition. *Psychol. Sci.* 19, 998–1006.
- Itier, R. J., and Batty, M. (2009). Neural bases of eye and gaze processing: the core of social cognition. *Neurosci. Biobehav. Rev.* 33, 843–863.
- Itier, R. J., Villate, C., and Ryan, J. D. (2007). Eyes always attract attention but gaze orienting is task-dependent: evidence from eye movement monitoring. *Neuropsychologia* 45, 1019–1028.

- Janik, S. W., Wellens, A. R., Goldberg, M. L., and Dell'Osso, L. F. (1978). Eyes as the center of focus in the visual examination of human faces. *Percept. Mot. Skills* 47, 857–858.
- Johansen, M. K., and Palmeri, T. J. (2002). Are there representational shifts during category learning? *Cogn. Psychol.* 45, 482–553.
- Just, M. A., and Carpenter, P. A. (1980). A theory of reading: from eye fixations to comprehension. *Psychol. Rev.* 87, 329–354.
- Kimchi, R., Amishav, R., and Sulitzeanu-Kenan, A. (2008). Gender differences in global-local perception? Evidence from orientation and shape judgments. *Acta Psychol. (Amst.)* 130, 64–71.
- Kirouac, G., and Doré, F. Y. (1985). Accuracy of the judgement of facial expression of emotions as a function of sex and level of education. *J. Nonverbal Behav.* 91, 3–7.
- Lamberts, K. (1998). The time course of categorization. *J. Exp. Psychol. Learn. Mem. Cogn.* 24, 695–711.
- Langdell, T. (1978). Recognition of faces: an approach to the study of autism. *J. Child Psychol. Psych.* 19, 255–268.
- Langton, S. R. H. (2000). The mutual influence of gaze and head orientation in the analysis of social attention direction. *Q. J. Exp. Psychol.* 53A, 825–845.
- Levin, D. T. (2000). Race as a visual feature: using visual search and perceptual discrimination tasks to understand face categories and the cross race recognition deficit. *J. Exp. Psychol. Gen.* 129, 559–574.
- Lewin, C., and Herlitz, A. (2002). Sex differences in face recognition-women's faces make the difference. *Brain Cogn.* 50, 121–128.
- Liu, C. H., and Ward, J. (2006). The use of 3D information in face recognition. *Vision Res.* 46, 768–773.
- Loftus, G. R., and Mackworth, N. H. (1978). Cognitive determinants of fixation location during picture viewing. *J. Exp. Psychol. Hum. Percept. Perform.* 4, 565–572.
- Looser, C. E., and Wheatley, T. (2010). The tipping point of animacy: how, when, and where we perceive life in a face. *Psychol. Sci.* 21, 1854–1862.
- MacDorman, K. (2005). "Androids as an experimental apparatus: why is there an uncanny valley and can we exploit it?," in *Toward Social Mechanisms of Android Science: A CogSci 2005 Workshop*, Stresa, 106–118.
- MacDorman, K. F. (2006). "Subjective ratings of robot video clips for human likeness, familiarity, and eeriness: an exploration of the uncanny valley [electronic resource]," in *ICCS/CogSci-2006 Long Symposium: Toward Social Mechanisms of Android Science*. Vancouver.
- MacDorman, K., Green, R., Ho, C.-C., and Koch, C. (2009a). Too real for comfort? Uncanny responses to computer generated faces. *Comput. Hum. Behav.* 25, 695–710.
- MacDorman, K. F., Vasudevan, S. K., and Ho, C.-C. (2009b). Does Japan really have robot mania? Comparing attitudes by implicit and explicit measures. *AI Soc.* 23, 485–510.
- McBain, R., Norton, D., and Chen, Y. (2009). Females excel at basic face perception. *Acta Psychol. (Amst.)* 130, 168–173.
- McGivern, R. F., Mutter, K. L., Anderson, J., Wideman, G., Bodnar, M., and Huston, P. J. (1998). Gender differences in incidental learning and visual recognition memory: support for a sex difference in unconscious environmental awareness. *Pers. Individ. Dif.* 25, 223–232.

- Minato, T., Shimada, M., Itakura, S., Lee, K., and Ishiguro, H. (2006). Evaluating human likeness of an android by comparing gaze behaviors elicited. *Adv. Robot.* 20, 1147–1163.
- Minut, S., Mahadevan, S., Henderson, J. M., and Dyer, F. C. (2000). "Face recognition using foveal vision," in *Biologically Motivated Computer Vision Lecture Notes in Computer Science*, eds S.-W. Lee, H. H. Bülthoff, and T. Poggio (Berlin: Springer), 424–433.
- Miyahira, A., Morita, K., Yamaguchi, H., Morita, Y., and Maeda, H. (2000a). Gender differences and reproducibility in exploratory eye movements of normal subjects. *Psychiatry Clin. Neurosci.* 54, 31–36.
- Miyahira, A., Morita, K., Yamaguchi, H., Nonaka, K., and Maeda, H. (2000b). Gender differences of exploratory eye movements: a life span study. *Life Sci.* 68, 569–577.
- Montagne, B., Kessels, R. P., Frigerio, E., de Haan, E. H., and Perrett, D. I. (2005). Sex differences in the perception of affective facial expressions: do men really lack sensitivity? *Cogn. Process.* 6, 136–141.
- Mori, M. (1970). "Bukimi no tani: the uncanny valley. *Energy* 7(4):33–35 (in Japanese)," in *English translation provided at CogSci '05 workshop: toward social mechanisms of android science, views of the Uncanny Valley*. Available at: <http://www.androidscience.com/theuncannyvalley/proceedings2005/uncannyvalley.html>.
- Nowicki, S. Jr., and Hartigan, M. (1988). Accuracy of facial affect recognition as a function of locus of control orientation and anticipated interpersonal interaction. *J. Soc. Psychol.* 128, 363–372.
- Oliva, A., and Schyns, P. G. (1997). Coarse blobs or fine edges? Evidence that information diagnosticity changes the perception of complex visual stimuli. *Cogn. Psychol.* 34, 72–107.
- Pearson, A., Henderson, J., Schyns, P. G., and Gosselin, F. (2003). Task dependent eye movements during face recognition. *Abstr. Psychon. Soc.* 8, 84.
- Pollick, F. E. (2010). "In search of the uncanny valley," in *User Centric Media*, eds P. Daras, and O. M. Ibarra (Berlin: Springer), 69–78.
- Pomianowska, I., Germeys, F., Verfaillie, K., and Newell, F. N. (2011). The role of social cues in the deployment of spatial attention: head-body relationships automatically activate directional spatial codes in a Simon task. *Front. Integr. Neurosci.* 6:4.
- Posner, M. I. (1980). Orienting of attention. *Q. J. Exp. Psychol.* 32, 3–25.
- Ramey, C. H. (2005). The uncanny valley of similarities concerning abortion, baldness, heaps of sand, and humanlike robots. *Paper Presented at the Proceedings of Views of the Uncanny Valley Workshop: IEEE-RAS International Conference on Humanoid Robots*, Tsukuba.
- Rayner, K. (1987). Eye movements in reading and information processing. *Psychol. Bull.* 85, 618–660.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychol. Bull.* 124, 372–422.
- Rayner, K., and Fischer, D. L. (1987). Letter processing during eye fixations in visual search. *Percept. Psychophys.* 42, 87–100.
- Rehman, J., and Herlitz, A. (2007). Women remember more faces than men do. *Acta Psychol. (Amst.)* 124, 344–355.
- Remington, R. W., and Folk, C. L. (2001). A dissociation between attention and selection. *Psychol. Sci.* 12, 511–515.

- Robinson, D. A. (1964). The mechanics of human saccadic eye movement. *J. Physiol.* 174, 245–264.
- Saether, L., Belle, W. V., Laeng, B., Brennen, T., and Overvoll, M. (2009). Anchoring gaze when categorizing faces' sex: evidence from eye-tracking data. *Vision Res.* 49, 2870–2880.
- Saygin, A. P., Chaminade, T., Ishiguro, H., Driver, J., and Frith, C. (2012). The thing that should not be: predictive coding and the uncanny valley in perceiving human and humanoid robot actions. *Soc. Cogn. Affect. Neurosci.* 7, 413–422.
- Scholten, M. R., Aleman, A., Montagne, B., and Kahn, R. S. (2005). Schizophrenia and processing of facial emotions: sex matters. *Schizophr. Res.* 78, 61–67.
- Schwarzer, G., Huber, S., and Dümmler, T. (2005). Gaze behavior in analytical and holistic face processing. *Mem. Cognit.* 33, 344–354.
- Schyns, P. G. (1991). A modular neural network of concept acquisition. *Cogn. Sci.* 15, 461–508.
- Schyns, P. G., Bonnar, L., and Gosselin, F. (2002). Show me the features! Understanding recognition from the use of visual information. *Psychol. Sci.* 13, 402–409.
- Schyns, P. G., and Gosselin, F. (2003). “Diagnostic use of scale information for componential and holistic recognition,” in *Perception of Faces, Objects, and Scenes. Analytic and Holistic Processes*, eds M. A. Peterson, and G. Rhodes (Oxford: Oxford University Press), 120–145.
- Schyns, P. G., and Murphy, G. L. (1994). “The ontogeny of part representation in object concepts,” in *The Psychology of Learning and Motivation*, ed. D. L. Medin (New York: Academic Press), 305–349.
- Schyns, P. G., and Oliva, A. (1994). From blobs to boundary edges: evidence for time and spatial scale dependent scene recognition. *Psychol. Sci.* 5, 195–200.
- Sergent, J. (1986). “Microgenesis of face perception,” in *Aspects of Face Processing*, eds H. D. Ellis, M. A. Jeeves, F. Newcombe, and A. M. Young (Dordrecht: Martinus Nijhoff), 17–33.
- Seyama, J., and Nagayama, R. S. (2007). The uncanny valley: the effect of realism on the impression of artificial human faces. *Presence (Camb.)* 16, 337–351.
- Shen, J., Reingold, E. M., Pomplun, M., and Williams, D. E. (2003). “Saccadic selectivity during visual search: the influence of central processing difficulty,” in *The Mind's Eyes: Cognitive and Applied Aspects of Eye Movement Research*, eds J. Hyona, R. Radach, and H. Deubel (Amsterdam: Elsevier Science Publishers), 65–88.
- Stacey, P. C., Walker, S., and Underwood, J. D. (2005). Face processing and familiarity: evidence from eye-movement data. *Br. J. Psychol.* 96(Pt 4), 407–422.
- Tanaka, J. W., and Curran, T. (2001). A neural basis for expert object recognition. *Psychol. Sci.* 12, 43–47.
- Tavassoli, A., Linde, I., Bovik, A. C., and Cormack, L. K. (2009). Eye movements selective for spatial frequency and orientation during active visual search. *Vision Res.* 49, 173–181.
- Thayer, J. F., and Johnsen, B. H. (2000). Sex differences in judgement of facial affect: a multivariate analysis of recognition errors. *Scand. J. Psychol.* 41, 243–246.
- Tinwell, A., and Grimshaw, M. (2009). “Bridging the uncanny: an impossible traverse?,” in *Proceedings of the 13th International MindTrek Conference: Everyday Life in the Ubiquitous Era*, eds O. Sotamaa, A. Lugmayr, H. Franssila, P. Näränen, and J. Vanhala (Tampere: ACM), 66–73.

- Tinwell, A., Grimshaw, M., and Williams, A. (2010). Uncanny behaviour in survival horror games. *J. Gaming Virtual World*. 2, 3–25.
- Tyler, C. W., and Chen, C. C. (2006). Spatial summation of face information. *J. Vis.* 6, 1117–1125.
- van Belle, G., Ramon, M., Lefèvre, P., and Rossion, B. (2010). Fixation patterns during recognition of personally familiar and unfamiliar faces. *Front. Psychol.* 1:20.
- Vuong, Q. C., Peissig, J. J., Harrison, M. C., and Tarr, M. J. (2005). The role of surface pigmentation for recognition revealed by contrast reversal in faces and Greebles. *Vision Res.* 45, 1213–1223.
- Walker-Smith, G. J., Gale, A. G., and Findlay, J. M. (1977). Eye-movement strategies involved in face perception. *Perception* 6, 313–326.
- Walters, M. L., Syrdal, D. S., Dautenhaun, K., Boekhorst, R., and Koay, K. L. (2008). Avoiding the uncanny valley: robot appearance, personality and consistency of behavior in an attention-seeking home scenario for a robot companion. *Auton. Robots* 24, 159–178.
- Yarbus, A. L. (1967). *Eye movements and Vision*. New York: Plenum Press.

5. Study C. Perceptual discrimination difficulty and familiarity in the Uncanny Valley

5.1. Abstract

The *Uncanny Valley Hypothesis* predicts that enhanced difficulty perceptually discriminating between categorically ambiguous human and humanlike characters (e.g., highly realistic robot) will evoke negatively valenced (i.e., uncanny) experience. An ABX perceptual discrimination task and signal detection analyses were used to delineate the profile of *perceptual discrimination (PD)* difficulty along the UVH' *dimension of human likeness (DHL)*. The DHL was represented using avatar-to-human morph continua. Rejecting the implicit assumption underlying the UVH' prediction, Experiment 1 (N=49) confirmed reduced PD difficulty for categorically ambiguous (and for unambiguous avatar) faces and, notably, enhanced PD difficulty for unambiguous human faces. In contrast to the predicted direction of the relationship between PD difficulty and affective experience (i.e. the UVH' *familiarity* dimension), Experiment 2 (N=49) demonstrated that greater PD difficulty is associated with more positively rather than negatively valenced affective experience. Critically, this effect was strongest for ambiguous faces, this suggesting a relationship between PD and familiarity that is more consistent with the metaphor *happy valley*. Using inverted faces, Experiment 3 (N=25) found no evidence to suggest that the asymmetry in PD along the DHL is attributable to a differential processing strategy (i.e., processing avatars at a category and human faces at an individual level). In conclusion, these findings for static faces show clearly that the UVH' assumed distribution of PD difficulty along the DHL and the predicted relationship between this and familiarity is incorrect.

5.2. Introduction

Progress in robotics and computer graphics in simulating human appearance and behaviour to high degrees of realism has fuelled research interest in the *Uncanny Valley Hypothesis (UVH)* (Mori, 1970). The UVH predicts that perceptual difficulty discriminating between highly realistic humanlike objects and characters (e.g., robot, prosthetic hand) and their human equivalent will evoke an unpleasant affective state.

The degree of high realism at which this state occurs is conjectured to be at the point along the UVH' *dimension of human likeness (DHL)* at which attribution of objects and characters to the human or nonhuman category is subject to greatest ambiguity (i.e., the 'valley' in Fig.1). This unpleasant state is described as marked by a sense of personal disquiet, strangeness and the uncanny. While the data have not yet provided a consistent picture in favor of this uncanny effect, this field of research is still in its infancy (e.g., Hanson, 2006; MacDorman, 2006). Possibly for this reason, almost no attention has been given to where along the DHL there is greater difficulty in *perceptual discrimination (PD)* and whether *PD* difficulty does relate to affective experience as described in the UVH (for exceptions, see Cheetham et al., 2011; Looser and Wheatley, 2011).

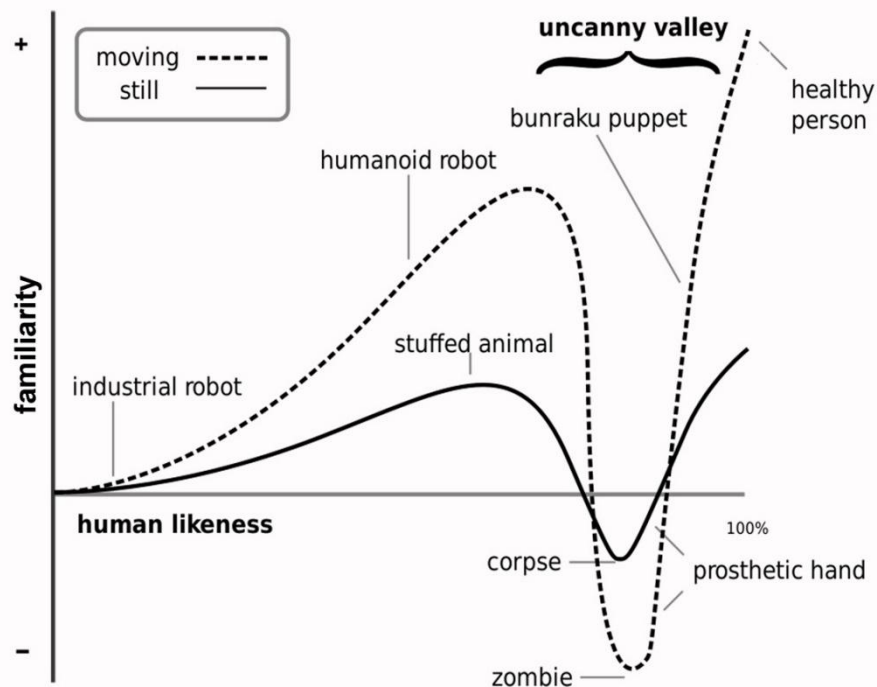


Figure 1. Illustration of the Uncanny Valley Hypothesis. The uncanny valley hypothesis proposes a non-linear relationship between affective experience and physical humanlike realism in appearance and motion (the non-linearity is more pronounced for motion). The key prediction of the hypothesis is that a high degree of human likeness will evoke a sharp negative peak (valley) in affective experience (i.e., along the familiarity dimension). This valley is characterized by feelings of strangeness (and the uncanny). The valley occurs at the point along the dimension of human likeness at which objects are categorically most ambiguous (illustration from MacDorman, 2005).

The UVH' prediction is based on the implicit assumption that PD difficulty is enhanced at or near the point of greatest category ambiguity (i.e., the category boundary) along the DHL. This assumption conflicts with the general consensus that lesser rather than greater difficulty in PD is normally associated with a category boundary (e.g., Cheetham et al., 2011). Besides this problem, a further potential problem that was not addressed in recent theoretical work on the uncanny effect (Moore, 2012) is that there might be asymmetry in PD difficulty along the DHL such that PD difficulty is actually greatest for categorically unambiguous human stimuli. If this is the case, and assuming that the conjectured direction of the relationship between greater PD difficulty and more negative affective experience is correct, it would follow from the UVH' prediction that greater negative affect should actually be experienced for objects and characters at the human end of the DHL. This potential outcome of the UVH' prediction would be difficult to reconcile with the available evidence. This indicates more positive affective experience for the human compared with nonhuman stimuli along the DHL (e.g. XXXX), unless of course the direction of the conjectured relationship between PD difficulty and affect is also incorrect.

These two problems can be considered in terms of the literature on *categorical perception (CP)*. CP refers to the phenomenon that the cognitive representation of *psychological similarity space* (such as along a perceptual dimension like the DHL) can be selectively deformed (Livingstone et al., 1998). Deformation (or warping) of psychological similarity space is evident when, relative to a baseline of comparison, physical differences between stimuli within a category are subjectively perceived to be more similar (i.e., less discriminable) than equally spaced physical differences between stimuli from two different categories that are subjectively perceived to be less similar (i.e., more discriminable). Many studies of CP such as for famous faces (Beale and Keil, 1995), unfamiliar faces (Levin and Beale 2000), facial expressions (Ectoff and Magee 1992) and faces of different gender (Bülthoff and Newell, 2000) show a relatively symmetrical pattern of perceptual discrimination performance along the continua, with enhanced PD performance at the category boundary (rather than poor PD performance as assumed in the UVH) and attenuated PD performance within each category.

Studies of CP (including those in the preceding) typically assume comparable (or symmetrical) category knowledge, categorization experience and processing of continua endpoints from which morph continua are generated. In contrast, the UVH was originally formulated on the basis of (informal) observation of individuals with little perceptual and categorization experience of humanlike robotic characters. The implicit assumption in the UVH that categorization experience is asymmetrical is also implicit in most uncanny-related studies (e.g., Yamada et al., 2013; Burleigh et al., 2013). These typically examine participants who have everyday expertise in facial processing of human category exemplars (see Tanaka, 2001; Diamond & Carey, 1977) but, by virtue of the innovative nature of avatar and robot research and design, comparatively little if any such experience with the subtle perceptual manipulations of and differences in human likeness in the nonhuman objects and characters under investigation. It is conceivable that these differences in categorization experience might influence perceptual information processing, discriminative performance (Goldstone, 1994; Sigala et al., 2010; Harnad, 1987; Gibson, 1991; Hall, 1991) and category processing along the DHL (see the neuroimaging experiment in Cheetham et al., 2011). Compared with discrimination performance before categorization training, categorization experience with novel continua based on line drawings of fictitious animals (Livingston et al., 1998), natural unfamiliar faces (Kikutani et al., 2008; 2010) and faces of identical twins (Stevenage, 1998) is reflected in a reduction in the discriminability of within-category stimuli and an improvement in the discriminability of between category stimuli. It would be consistent with such findings that asymmetry in perceptual and categorization experience with human compared with novel nonhuman objects and characters might be reflected in a corresponding asymmetry in PD performance along the DHL. This would mean poorer PD performance for within-category human stimuli (for which healthy participants have everyday expertise due to a history of normal social interaction) compared with better PD performance for within-category nonhuman stimuli (for which participants have comparatively little or no such everyday expertise).

In the first of three experiments, we tested whether the distribution of PD difficulty along the DHL implicitly assumed in the UVH is correct. Considering the influence of categorization experience on CP and psychological similarity space reported in the preceding studies, we anticipated, firstly, that faces within the human category would

generally be more difficult to discriminate compared with those closest to or at the category boundary. Second, we anticipated that faces within the human category would generally be more difficult to discriminate compared with those within the nonhuman category. To examine this, we delineated the profile of PD performance for morphed faces drawn from morph continua representing the DHL. The continua were generated from avatar (i.e., computer-generated characters) and human parent faces. The morphed faces were presented in an *ABX perceptual discrimination task* (Liberman et al, 1957; this task is described in detail in Section 2.1.3). Campbell et al. (1997) used the ABX task to investigate CP along other dimensions of human likeness and show that this task is sensitive to perceptual processing differences between human and nonhuman faces. Signal detection analysis was used to assess discrimination sensitivity. A *two-alternative forced choice categorization task* (described in detail in Section 2.1.3) was conducted after the ABX task in order to define the profile of categorisation ambiguity and the location of the category boundary along the continua. The second experiment replicated the findings of the first experiment. In the second experiment, we tested the UVH' predicted relationship between PD difficulty and negative affective experience. In the third experiment, we explored the possibility that the asymmetry in PD difficulty reported in Experiments 1 and 2 might be attributable to a *differential processing bias* such that different category and exemplar-related information is extracted from avatar and human faces, respectively.

5.3. Experiment 5: Perceptual discrimination and forced choice categorization.

5.3.1 Materials and Methods

Participants

Healthy adult volunteers (N = 49, 29 female, mean age 21.8 years; range 19 - 25 years) with no record of neurological or psychiatric illness and no current medication use were recruited for the study. All study participants were students of the University of Zurich, native or fluent speakers of Swiss or Standard German, and consistently right-handed, as assessed with self-rating scales (Annett, 1970). Each participant

confirmed after completion of the experiment having had no previous experience designing or modifying computer-generated characters for example in virtual reality (VR) role-playing games, second life, or VR environments or using such environments (e.g. for psychotherapy, rehabilitation, training, e-commerce or virtual reality-based research) and explicitly no previous experience (e.g., in video games, cinema) with the kind of highly humanlike characters presented in the current study, with similarly humanlike characters or with manipulations of human likeness as in such a study. At debriefing, one participant reported uncertainty about the correct use of the response buttons and 3 others about the meaning of the label "avatar" in the forced choice categorization task. Analyses with and without the data of these four participants had no impact on the pattern of findings. The findings are reported on the basis of the complete data set. Written informed consent was obtained before participation according to the guidelines of the Declaration of Helsinki. Each volunteer received 20 Swiss Francs for participation. The study and all procedures and consent forms were approved by the Ethics Committee of the University of Zurich.

Stimuli

Morph continua were generated to represent the DHL, using the software Fantamorph® (Abrosoft, <http://www.abrosoft.com>). Eight colour photographic images of natural faces and 8 colour images of avatar faces were used as parent faces to produce 8 morph continua. The selection of continua was based on previous pilot testing to ensure like performance across continua (i.e. same morph position of the category boundary and shape of the response function). Each continuum comprised 11 different morphed images from the avatar endpoint (number 1) to the human endpoint (number 11), each morph being separated by an increment of 10% in physical difference (see Fig, 2.B). All parent faces were male, indistinctive, presented with full face, frontal view, direct gaze, neutral expression and no other salient features such as facial hair and jewellery. Avatars were generated with the modeling suite Poser 7® (Smith Micro Software, <http://www.smithmicro.com>) for detailed adjustment of facial geometry and texture (e.g. age and configural cues) to closely match the corresponding human face. Matching aimed to minimize perception of biological motion due to quick successive presentation of morphs (Schultz and Pilz, 2009) and to ensure perception of faces in the two-step procedure of the ABX as

having the same identity. Adobe Photoshop 7.0® (<http://www.adobe.com>) was used for image editing. Before morphing, the external features of each parent face were masked with an elliptic form and black background (96 dpi and 560 x 650 pixels), and contrast levels, overall brightness and skin tone of the parent faces of each continuum were adjusted to match.

Design and procedure

All participants were tested individually by a research assistant blind to the purpose and hypotheses of the study. The perceptual discrimination ABX task was conducted first followed by the two-alternative forced-choice categorization task. The experiment lasted approximately 40 min, with a short break between the discrimination and categorization tasks.

Perceptual discrimination ABX task

The UVH does not suggest how DP difficulty should be operationalised and tested. For this, was used the ABX discrimination task (Liberman et al., 1957; Harnad, 1987). This entails presentation of trials in which pairs of different face stimuli (A and B) are followed by a second presentation of either A or B as the target stimulus X. Participants are required to view all three images and respond by button press to indicate whether A or B is identical to (i.e., the same as) X. A 2-step discrimination procedure was applied so that stimulus B differed in physical distance along the continuum from stimulus A by two steps (i.e., 1-3, 2-4, 3-5, etc.). To counterbalance the sequence of face pairs, each pair was presented four times, once in each of the possible combinations (i.e., AB-A, BA-B, AB-B, BA-A). Both faces of each presented pair were always drawn from the same continuum in which they were originally morphed. The presentation of face pairs was pseudo-randomized so that no trials using face pairs from corresponding morph positions of other continua were presented in sequence.

Written instructions were presented on the screen before commencement of the experiment. Participants performed a pre-test of 5 trials (using stimuli drawn at random from continua that were not included in the main test) to ensure

comprehension of the instructions and correct use of the response buttons. The background on the monitor was always black. Stimuli A and B were presented for 750ms immediately followed by stimulus X, which remained on screen until the response was made or till time-out at 4 s. The inter-trial interval was 1500ms. Response accuracy and *response time (RT)* were measured for each trial, including the practice trials.

The ABX (and forced-choice classification task described in the next section) were conducted in a sound attenuated and light-dimmed room, and morph stimuli were presented on a LCD monitor (1280 x 1024 resolution, 60 Hz refresh rate), using Presentation® software (Version 14.1, www.neurobs.com). The stimuli (400 × 500 pixels) were presented at a viewing distance of 62 cm.

Two-alternative forced-choice categorization task

The same stimuli presented in the ABX task were presented in a two-alternative forced-choice categorization task. This task commenced with the presentation of written instructions. Subsequently, participants performed a practice pre-test of 5 trials, using the same stimuli used in the pre-trials of the ABX task. Having ensured task comprehension and correct use of the response buttons, the participant initiated testing by pressing a button. The forced-choice categorization task normally follows the perceptual discrimination task in order to minimize the potential influence of labeling on discrimination performance (Newell & Bulthoff, 2002). To minimize this further, the labels "avatar" and "human" were first used during task instruction for the forced choice task. The background on the monitor was always black. All morph stimuli were presented twice, individually, centrally, and in random order with the constraint that stimuli from corresponding morph positions of other continua were not presented in sequence. Each trial began with the presentation of a fixation point for 500ms (participants were required to maintain fixation), followed by a morph image for 750ms. The participant was asked to identify the stimulus quickly and accurately as either an avatar or human by pressing one of two response keys. A black screen with fixation point remained after presentation of the morph image until the participant pressed the response key, after which a blank black screen without fixation cross remained for 1500 ms until the next trial began.

All data analyses were performed using SPSS version 21.0 (<http://www.ibm.com>). MATLAB 2006b (<http://www.mathworks.ch>) was used to implement the Palamedes routines (Prins and Kingdom, 2009) for signal detection analysis of data from the ABX task.

5.3.2. Results

The response data for avatar-versus-human category judgments in the forced choice categorization task were analysed (Section 2.2.1) to determine the choice of categorically ambiguous and unambiguous morphs for use in the analyses of perceptual discrimination performance (Section 2.2.2).

Forced choice categorization task: Responses, logistic function, and category boundary

The slope of the categorization response function was used to summarize the category judgments by fitting logistic function models to the data of each participant across continua. The parameter estimates derived from each model were entered in analyses of logistic function of categorization responses and of the category boundary.

For the logistic function of categorization responses, the parameter estimates were tested against zero in a one-sample t-test. The result showed a highly significant logistic component ($t_{48} = 44.31$, $p > 0.001$) consistent with the presence of a category boundary (Harnad, 1987) (see Fig. 2).

To compute the value of the category boundary (i.e., $y=.5$: $-\ln[\beta_0]/\ln[\beta_1]$), the parameter estimates β_0 and β_1 of each participant's logistic function model were used. The mean category boundary value was $M = 6.95$. This value indicates the actual morph position along the continua that corresponds with the ordinate midpoint between the lower and upper asymptotes, that is, the point of maximum uncertainty of 50% in categorization judgments. Across continua, morph M7 is closest to this boundary (Fig. 2).

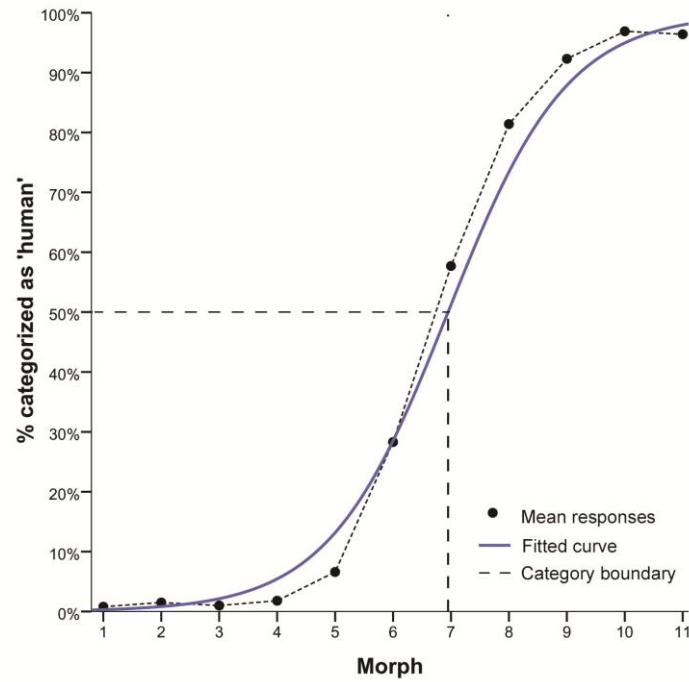


Figure 2. Results of the forced choice classification task. Mean responses are depicted in terms of % of ‘human’ responses (A). The mean grand average (continuous blue line) across all continua (error bars showing standard deviation), fitted logistic curve based on the grand mean (black line), and the category boundary (dashed gray line) are shown. The category boundary indicates the point of maximum uncertainty of 50% in categorisation judgements along the continua. The logistic-shaped curve shows a lower and upper asymptote of avatar and human categorisation responses and a step-like response function consistent with the presence of a category boundary. Morph M7 shows the greatest categorisation ambiguity.

To show this profile of high and low ambiguity in categorization judgments more clearly, we tested for differences in category decisions between the unambiguous avatar (i.e., M3, M4, M5) and human faces (i.e., M9, M10, M11) and the most ambiguous faces (i.e., M7). This choice of morphs permitted control for physical morph distance along continua between the ambiguous M7 and the unambiguous avatar and human faces. A one-way *repeated measures of analysis of variance (RM-ANOVA)* was performed on the dependent variable mean ‘categorization’ response of each participant across continua, using the factor ‘morph’ position (3 levels: “M3, M4, M5”, “M7”, “M9, M10, M11”). Greenhouse-Geisser adjustment was applied to correct the degrees of freedom for violation of the sphericity assumption (and applied as appropriate in all subsequent analyses). This analysis showed a highly significant effect for morph position, $F(1.27, 58.93) = 455.26, p < 0.001$. Mean categorization

difficulty for M7 was $M = 0.58$ ($SE = 0.04$), while that for the human faces was $M = 94.52$ ($SE = 0.01$) and for avatar faces $M = 4.02$ ($SE = 0.01$) (see, Fig. 3).

Forced choice categorization task: Response times

Differences in category ambiguity, as indicated by the logistic-shaped response function, are likely to be reflected in different *response times (RT)* for category judgments. Before data analysis, short RT latencies of less than 100 ms were excluded. RT data for long latency outliers were screened by z-standardizing and filtering out data points using $z=3$ as a cut-off score (Van Selst and Jolicoeur, 1994). Analyses were conducted with and without outliers. These analyses produced the same pattern of results. The findings are therefore reported for the complete data set. Confirming RT differences in category decision difficulty, a one-way RM-ANOVA with morph position (11 levels: M1- M11) and RT as the dependent variable showed a main effect for morph position, $F(4.58, 215.09) = 41.23$, $p < 0.001$.

The longest response latencies would be expected to correspond with the morph position closest to the category boundary, that is, at M7. But inspection of the Figure X indicates that RT for M6 and M7 are similarly long. The tests of planned within-subject contrasts in the preceding analysis showed no significant difference between M6 and M7 in RT. Given that M7 and the category boundary are so closely aligned, the following analysis compared ambiguity at M7 with the unambiguous avatar (i.e., M3, M4, M5) and human faces (i.e., M9, M10, M11), but a re-run of the same analysis using the aggregate mean of M6 and M7 instead of just M7 produced the same pattern of results. A one-way RM-ANOVA analysis with 'morph' positions (3 levels: "M3, M4, M5", "M7", "M9, M10, M11") and RT in ms as dependent variable was conducted. The analysis showed a highly significant effect for morph position [$F(2.96, 58.93) = 45.22$, $p < 0.001$] (see Fig XX). Pre-planned contrasts showed that RT was longer significantly longer for human ($M = 1073$, $SE = 44$) than for avatar faces ($M = 898$, $SE = 33$), $F(1,48) = 15.72$, $p > 0.001$, and that RT for M7 ($M = 1348$, $SD = 60$) differed highly significantly from RT for the other avatar and human morph positions ($M = 928$, $SD = 0.19$), $F(1,48) = 67.49$, $p < 0.001$.

ABX Perceptual discrimination task

Differences in the ability to perceptually discriminate between pairs of morphs (M6-M8) straddling the ambiguous M7 and between pairs of unambiguous morphs within the avatar (M3-M5, M4-M6) and human (M8-M10, M9-M11) face categories were tested. This choice of avatar and human morph pairs ensured control for the physical morph distance along the continua between the ambiguous and unambiguous faces. The mean value of PD was compared in a one-way RM-ANOVA with factor morph position (3 levels: "M3-M5, M4-M6", "M6-M8", "M8-M10, M9-M11") using d' as dependent variable (Best et al., 1981). d' is used as a measure of discrimination performance derived from Signal Detection Theory (e.g., Macmillan and Creelman, 2005) that takes effects of response bias (c) into account. This measure is used instead of the percentage of correct different responses to different pairs (Francis and Ciocca, 2003). A differencing model was applied to compute d' because this is considered to best reflect the decision strategy used in the ABX task (Hautus and Meng, 2001; Macmillan and Creelman, 2005; Pierce and Gilbert, 1958).

This analysis showed a significant effect for morph pair position, $F(2, 96) = 14.68$, $p < 0.001$. Tests of planned within-subject contrasts showed that PD of faces within the avatar category ($M = 1.74$, $SE = 0.1$) was significantly greater than that of ambiguous faces at the category boundary ($M = 1.47$, $SE = 0.12$) ($F[1, 48] = 5.59$, $p = 0.022$) and of faces within the human category ($M = 1.15$, $SE = 0.05$), $F(1, 48) = 38.54$, $p < 0.001$. PD of ambiguous faces was significantly greater than that of faces within the human category, $F(1, 48) = 7.5$, $p = 0.009$.

A one-way RM-ANOVA with "morph position" (11 levels) and c as the dependent variable for response bias showed no significant differences for c .

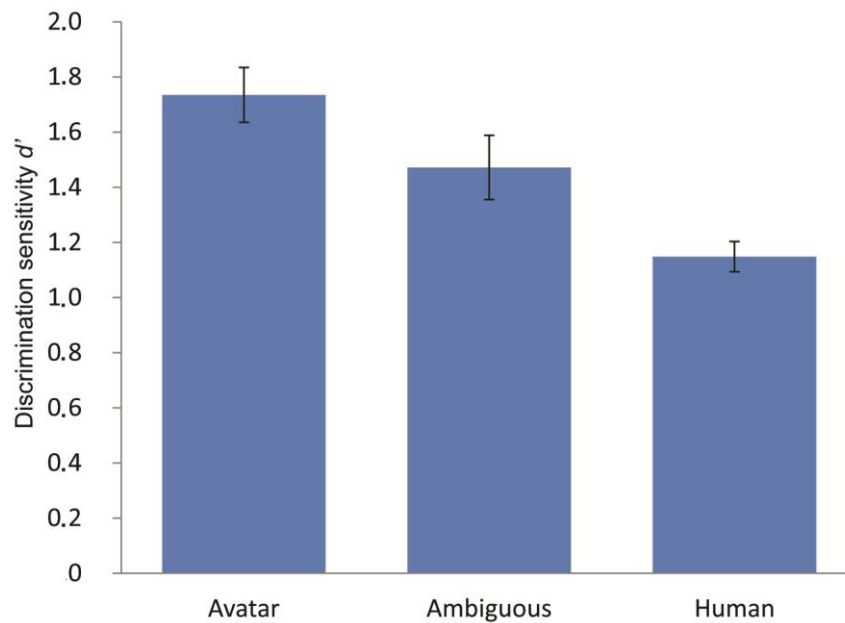


Figure 3. Results of the ABX perceptual discrimination task. The figure depicts mean discrimination sensitivity d' in the ABX perceptual discrimination task for unambiguous avatar and human and highly ambiguous faces. The profile of d' shows a marked asymmetry along the continua, meaning that unambiguous avatar (and the ambiguous faces) are perceived as more dissimilar than equally spaced human faces. Contrary to the implicit assumption in the UVH, perceptual discrimination difficulty is thus greatest for human faces. The error bars indicate 1 SE ($N = 49$).

5.3.3. Discussion

The data confirm that there are differences in PD difficulty as a function of human likeness along the DHL. But the pattern of PD is entirely different than that implicitly assumed in the UVH. Firstly, and as expected on the basis of previous studies of CP, PD of faces at the category boundary is enhanced compared with PD of within-category human faces. Second, PD of within-category avatars is also enhanced compared with PD of within-category human faces, thus supporting the suggestion that there might be asymmetrical PD along the DHL.

Given that the UVH predicts enhanced negative affective experience as a function of enhanced PD difficulty, these findings would mean - assuming that the UVH is otherwise correct - that human faces should evoke more negative affect compared with ambiguous faces and unambiguous avatar faces. This is clearly inconsistent with

the idea that Mori sought to convey in his graphical representation of his hypothesis, and the available evidence from uncanny-related research suggests that enhanced feelings of strangeness for human category exemplars is highly unlikely. Self-ratings of comparably well-controlled morph continua show that positive ratings (e.g., liking, pleasantness) increase with greater human likeness (e.g., Looser and Wheatley, 2011). More recently, well-validated psychophysiological and self-assessment measures of valence and arousal (i.e., the underlying orthogonal dimensions of affective experience) that are unaffected by the conceptual ambiguities surrounding the affective dimension of the UVH also demonstrate that positive affective experience increases with increasing human likeness (Cheetham et al., submitted in Frontiers).

In a second experiment, we tested whether there is nevertheless evidence in favor of the UVH' prediction that enhanced PD difficulty is associated with greater negative affective experience. The UVH conceptualizes affective experience as *shinwakan*, an ambiguous Japanese neologism that Mori used to describe the positive and negative character of affective experience of humanlike objects. There have been various renderings of shinwakan's meaning in uncanny-related research, including comfort level, familiarity, eeriness, pleasantness, likability, empathy and affinity (e.g. Burleigh et al., 2013; MacDorman et al., 2013; Dill et al., 2012; Mori, 2012; Tinwell et al., 2011; Seyama and Nagayama, 2009; Green et al., 2008; 2007; Bartneck et al., 2007; MacDorman and Ishiguro, 2006). To examine affective experience, we used an *ad hoc* self-rating scale based on the UVH' bi-polar dimension of *familiarity* (i.e., familiar vs. strange). Familiarity was selected because this rendering of shinwakan has been used frequently in research, it is most often used to denote the affective dimension of the UVH in its illustration (see Fig.1), and because it arguably best captures the intended meaning of shinwakan that Mori sought to convey in the UVH's description. Clearly, there are alternative approaches to examining affective experience of human like objects and characters based on well-validated dimensions of affective experience and measures of these (see Cheetham et al., in Frontiers). The aim of this experiment was to test affective experience as conceptualized in the UVH in relation to PD difficulty.

5.4. Experiment 6. Perceptual discrimination and familiarity.

5.4.1. Materials and Methods

The materials, methods and analyses in Experiment 2 were identical to those in Experiment 1, with two exceptions. Firstly, the presented morphs were drawn from continua that were generated anew by switching the source image (i.e., avatar) and destination image (i.e., human) for morphing in Experiment 1 so that the human was the source and the avatar the destination image. The continua were then re-morphed, and the morphs were labeled M1 (avatar) to M11 (human) as in Experiment 1. The aim of re-morphing was to exclude the possibility that the strong asymmetry in perceptual discrimination performance in Experiment 1 was simply a systematic artefact of any nonlinearity in the morphing algorithm. If it was a systematic artefact, the PD data in Experiment 2 would show a similar pattern of PD along the DHL skewed however toward enhanced PD for human instead of avatar faces. Second, participants performed the ABX task followed this time by a self-rating task, in which to report feelings of familiarity, and only then by the two-alternative forced choice categorization task. The latter task was performed last to ensure that any effects in ratings were not biased by explicit processing of faces for forced categorization.

The UVH does not suggest how DP difficulty and feelings of familiarity should be operationalised and tested. We used our measure of discrimination sensitivity d' to indicate DP performance, as applied in Experiment 1, and, in keeping with the favored approach to date in uncanny research, subjective ratings indicating feelings of familiarity in the self-rating task. The task requirements, instructions and stimulus presentation conditions for the self-rating task were identical to those described for the two-alternative forced choice categorization task in Experiment 1, with the exception that participants viewed and rated the subjective feeling of familiarity evoked by each morphed stimulus on a 5-point Likert scale. The scale ranged from very strange (1) to very familiar (5). To test the relationship between DP difficulty and feelings of familiarity, we took an inter-individual differences approach. We tested

whether individual variability in the ability to discriminate between a pair of morphed faces (e.g., M2-M4) predicts individual variability in self-rated feelings of familiarity for the face (e.g., M3) that the given face pair straddles. This approach assumes that there are stable individual differences in the relationship between familiarity ratings and discrimination performance. If Mori's prediction is correct, greater PD difficulty should be associated with increased feelings of strangeness (i.e., with less familiarity). This was tested.

Participants

A new sample of $N = 49$ volunteers (34 female, mean age 21.9 years; range 19 - 31 years) not involved in Experiment 1 participated in Experiment 2.

5.4.2. Results

Forced choice categorization task: Responses, logistic function, and category boundary

The parameter estimates derived from each logistic function model of each participant across continua were tested against zero in a one-sample t-test and showed, as in Experiment 1, a highly significant logistic component ($t_{48} = 27.83$, $p > 0.001$ (see Fig. 4). Based on the parameter estimates β_0 and β_1 , the mean category boundary value was $M = 6.6$. Across continua, the most ambiguous face morph M6 is closest to this boundary.

To show the effects of this profile of high and low ambiguity in categorization judgments more clearly, we tested for differences in category decisions between the unambiguous avatar (i.e., M2, M3, M4) and human faces (i.e., M8, M9, M10) and the most ambiguous faces (i.e., M6). Consistent with the approach in Experiment 1, the choice of morphs permitted control for physical morph distance along continua between M6 at the category boundary and the avatar and human faces. A one-way *RM-ANOVA* was performed on the dependent variable mean 'categorization' response of each participant across continua, using the factor 'morph' position (3 levels: "M2, M3, M4", "M6", "M8, M9, M10"). This analysis showed a highly significant

effect for morph position [$F(1.22, 58.52) = 483.72, p < 0.001$]. Categorization difficulty for M6 was closest to chance level of 50% ($M = 40.31$; $SE = 3.73$), while that for the human faces was $M = 93.58$ ($SE = 1.13$) and for avatar faces $M = 2.63$ ($SE = .44$) (see, Fig. 4).

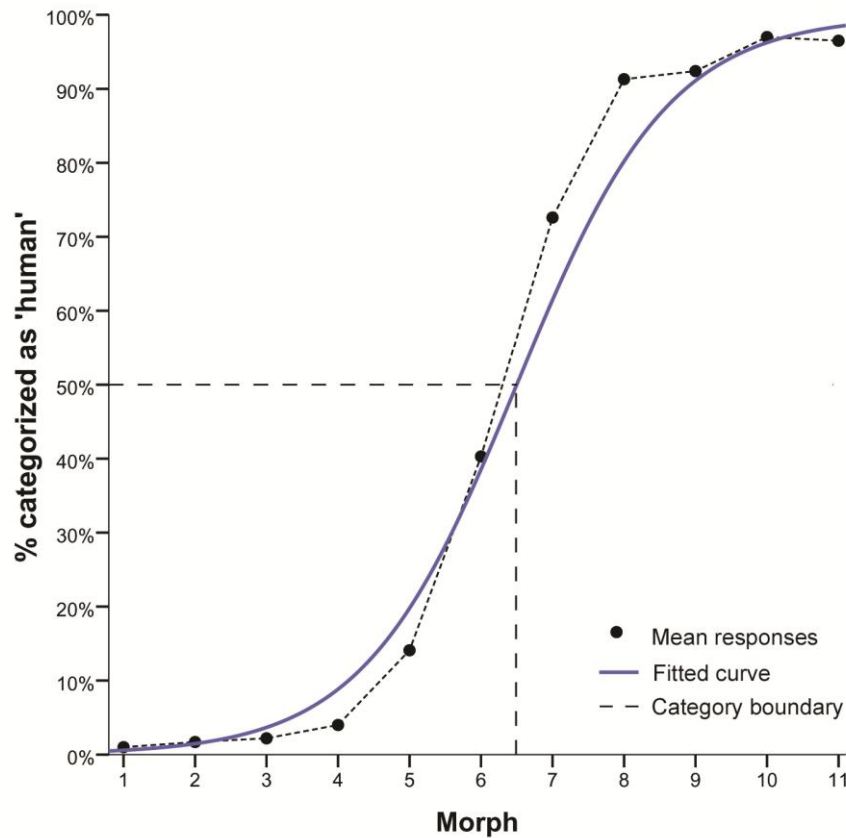


Figure 4. Results of the forced choice classification task. Mean responses are depicted in terms of % of “human” responses, with the grand average (continuous blue line) across all continua (with error bars showing standard deviation), the fitted logistic curve based on the grand mean (black line), and the category boundary (dashed gray line) to indicate the point of maximum uncertainty of 50% in categorisation judgements along the continua. Results indicate a step-like response function consistent with the presence of a category boundary. Morph M6 shows the greatest categorisation ambiguity.

Forced choice categorization task: Response times

We verified whether differences in category ambiguity are reflected in the RT for category judgments. Data were screened for outliers as in Experiment 1 and analyses conducted with and without these. These analyses produced the same

pattern of results for which reason the findings for the complete data set are reported. Confirming RT differences in category decision difficulty, a one-way RM-ANOVA with morph position (11 levels: M1-M11) and RT as the dependent variable showed a main effect for morph position, $F(4.58, 220.22) = 39.03, p < 0.001$.

Inspection of the RT data indicates that the longest response latencies correspond with the most ambiguous morph M6. A one-way RM-ANOVA analysis with 'morph' positions (3 levels: 'M2, M3, M4', 'M6', 'M8, M9, M10') and RT in ms as dependent variable was conducted. The analysis showed a highly significant effect for morph position, $[F(2, 96) = 54.99, p < 0.001]$. Pre-planned contrasts showed that RT was longer significantly longer for human ($M = 957, SE = 34$) than for avatar faces ($M = 751, SE = 23$), $F(1,48) = 15.72, p > 0.001$, and that RT for M6 ($M = 1191, SD = 53$) differed highly significantly from RT for the other morph positions ($M = 851, SD = 0.37$), $F(1,48) = 65.91, p < 0.001$.

ABX perceptual discrimination task

An independent samples t-test (Experiment 1 versus Experiment 2) using d' for each morph pair position in the ABX task (i.e., pairs M1-3 through to M9-11) of each participant across continua as dependent variable showed that discrimination performance for each morph pair was not significantly different between Experiments 1 and 2. The following results indicate also that the PD effects in Experiment 1 are comparable to those in Experiment 2.

Given that face morph position M6 was closest to the category boundary in Experiment 2, differences in the ability to perceptually discriminate between pairs of morphs (M5-M7) straddling the *ambiguous* M6 was compared with the ability to perceptually discriminate between unambiguous morphs within the *avatar* (M1-M3, M2-M4, M3-M5) and *human* (M7-M9, M8-M10, M9-M11) face categories. These morph pairs were selected because they straddle the morph positions M2, M3, M4, M6, M8, M9, M10 that were analysed in the forced choice task of Experiment 2 and because this choice of pairs ensures control for physical morph distance between the ambiguous and the unambiguous human and avatar faces. The mean value of discrimination sensitivity was compared in a one-way RM-ANOVA with factor morph

position (3 levels: "M1-M3, M2-M4, M3-M5", "M5-M7", " M7-M9, M8-M10, M9-M11") using d' as dependent variable.

This analysis showed a significant effect for morph pair position, $F(2, 96) = 16.52$, $p < 0.001$ (see, Fig. 5). Tests of planned within-subject contrasts showed that discrimination of avatar faces ($M = 1.41$, $SE = 0.08$) was significantly greater than that of faces within the human category ($M = 0.99$, $SE = 0.07$), $F(1, 48) = 27.59$, $p > 0.001$. Discrimination of ambiguous faces at the category boundary ($M = 1.53$, $SE = 0.11$) was not significantly greater than that of faces within the avatar category ($M = 1.41$, $SE = 0.08$) ($F[1, 48] = 1.11$, $p = 0.299$), but it was significantly greater than that of faces within the human category ($M = .99$, $SE = 0.07$), $F(1, 48) = 25.76$, $p < 0.001$.

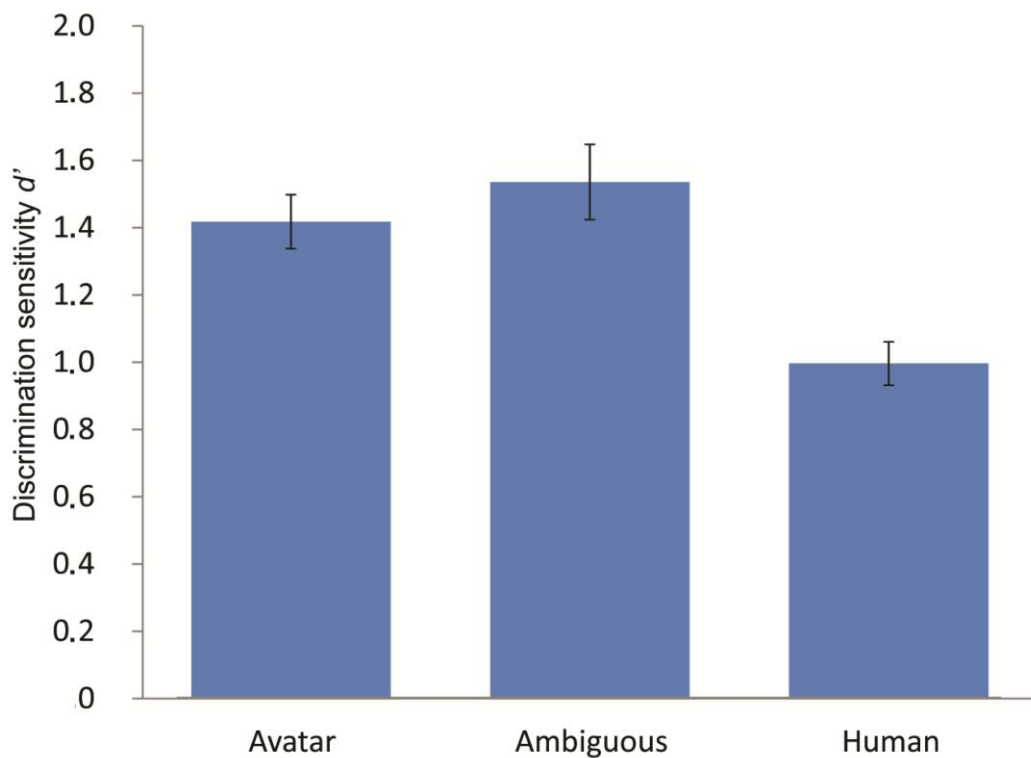


Figure 5. Results of the ABX perceptual discrimination task. The figure depicts mean discrimination sensitivity d' in the ABX perceptual discrimination task for unambiguous avatar and human and highly ambiguously faces. The profile of discrimination sensitivity replicates that found in the first experiment. Perceptual discrimination difficulty is greatest for human faces. The error bars indicate 1 SE ($N = 49$).

It should be noted that the most ambiguous morph was M7 in Experiment 1 and M6 in Experiment 2. This means that the choice of morph pairs for inclusion in the

analyses of d' in Experiment 1 is partially different than the choice in Experiment 2. To compare Experiments 1 and 2, the one-way RM-ANOVA in Experiment 2 was re-run, using this time the same morph positions selected in Experiment 1, that is, M3-M5 and M4-M6 for avatar faces, M6-M8 for the ambiguous M7, and M8-M10 and M9-M11 for human faces. This analysis showed the same pattern of significant effects for morph pair position [$F(2, 96) = 21.42, p < 0.001$] and for the tests of planned within-subject contrasts. The contrasts showed that PD of faces within the avatar category ($M = 1.67, SE = 0.11$) was significantly greater than that of ambiguous faces at the category boundary ($M = 1.34, SE = 0.11$) ($F(1, 48) = 12.87, p = 0.001$) and of faces within the human category ($M = 0.98, SE = 0.07$), $F(1, 48) = 35.48, p < 0.001$. PD of ambiguous faces was significantly greater than for faces within the human category, $F(1, 48) = 11.26, p = 0.002$. Taken together, these analyses are consistent in indicating asymmetry in discrimination performance along the continua.

A one-way RM ANOVA with "morph position" (11 levels) and c as the dependent variable for response bias showed no significant differences for c .

Familiarity ratings

Differences in mean familiarity ratings between the *unambiguous* avatar (i.e., M2, M3, M4) and *human* faces (i.e., M8, M9, M10) and the most *ambiguous* faces (i.e., M6) were tested using the same morph positions as in the analysis of the forced choice categorization task in Experiment 2 (Section 3.2.1). A one-way RM-ANOVA with the factor *morph* position (3 levels: 'M2, M3, M4', 'M6', 'M8, M9, M10') and the dependent variable *familiarity* rating of each participant across continua revealed a highly significant effect of morph position, $F(1.48, 70.93) = 180.61, p = < 0.001$ (see Fig.5). Pre-planned contrasts showed a significant difference between the avatar morphs ($M = 1.93; SE = 0.1$) and M6 ($M = 3; SE = 0.08$) [$F(1, 48) = 278.67, p = < 0.001$] and between M6 and the human morphs ($M = 3.68; SE = 0.07$), $F(1, 48) = 53.02, p = < 0.001$, and between the avatar and human morphs, $F(1, 48) = 27.59, p = < 0.001$. Taken together, the data indicate that familiarity ratings increase negatively (i.e., greater strangeness) across the three stimulus conditions with increasing distance from the human end of the continua. This lends no support to the UVH' predicted increase in negative evaluations for the most ambiguous faces.

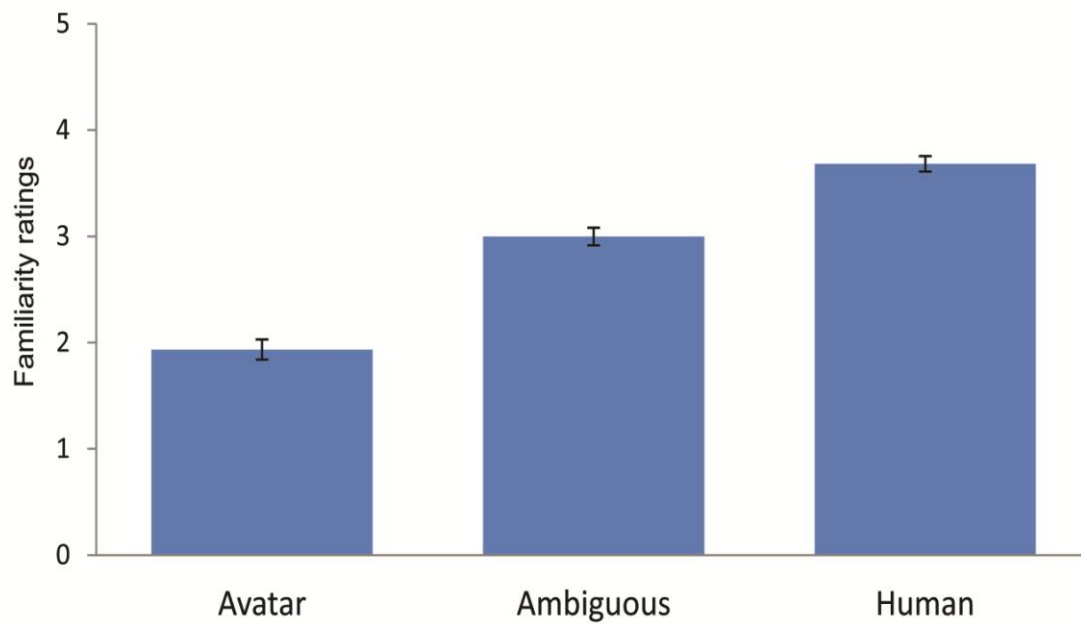


Figure 6. Results of self-rating task for familiarity. Overall, the figure illustrates a general increase in self-rated feelings of strangeness with decreasing human likeness from the human end of the continua. There is no indication of the uncanny effect predicted in the UVH. The error bars indicate standard errors (N = 49).

Relationship between perceptual discrimination and familiarity ratings

The UVH predicts a positive relationship between greater PD difficulty and greater subjective experience of strangeness. To test this we examined whether individual variability in PD for face pairs predicts individual variability in ratings of subjective experience for the faces that the face pairs straddle. Pearson product-moment correlations were conducted using the mean data of each participant across continua of each *morph* in the familiarity rating task (i.e. 'M2, M3, M4' for avatar, M6 for ambiguous, and 'M8, M9, M10' for human faces) and the morph pairs that straddled these faces in the ABX task (i.e. 'M1-M3, M2-M4, M3-M5' for *avatar*, M5-M7 for ambiguous, M7-M9, M8-M10, M9-M11 for human faces). Outlier detection was performed before analysis by means of boxplots. This indicated 1 outlier. After removal, the analyses showed a highly significant (two-sided) negative correlation between PD performance and familiarity ratings for avatar faces ($r_{48} = -.314$, $p = 0.03$) and for ambiguous faces ($r_{48} = -.494$, $p > 0.001$). There was no significant relationship

between PD performance and familiarity ratings for human faces ($r_{49} = .088$, $p = 0.533$).

5.4.3. Discussion

Experiment 2 replicated Experiment 1 by showing the same pattern of PD asymmetry, that is, enhanced perceptual discrimination for highly ambiguous faces and highly unambiguous nonhuman faces but attenuated discrimination for highly unambiguous human faces. Based on a new sample of participants and re-morphed continua, this pattern re-affirms that the implicit assumption in the UVH, that is, greater PD difficulty in the categorically most ambiguous region of the DHL, is incorrect. It is in this region that the UVH suggests stronger feelings of strangeness compared with those evoked by neighboring less ambiguous human or humanlike stimuli. But the data show that greater feelings of strangeness are actually reported for the least human faces, and that feelings of strangeness diminish with increasing human likeness of the facial morphs.

While there is no indication of an uncanny effect, these data are based on group averaging. It is however possible that there are inter-individual differences in the relationship between familiarity and PD difficulty and that these differences reveal an effect consistent with Mori's suggestion. In fact, the correlative data show a significant relationship between PD difficulty and feelings of familiarity, but the direction of this relationship is the opposite of that predicted in the UVH. Increasing PD difficulty is associated with more positive feelings of familiarity. Interestingly, this effect only applies for nonhuman and ambiguous faces. There was no significant relationship between PD difficulty and familiarity for human faces. Critically, this correlative effect was greatest for ambiguous faces. Taken together, the correlative data suggest, irrespective of the question of the causal direction, that the UVH' prediction is most likely wrong.

The reason for asymmetry in perceptual discrimination performance along the continua is not clear. One potential explanation draws on the suggestion that human observers preferentially code other members of the human in-group (e.g., our human exemplars) differently than members of a nonhuman out-group (e.g., our highly

humanlike avatars) (Cheetham et al., 2013; see the *other-race hypothesis*, Levin, 2000; *differential processing hypothesis*, Ostrom et al., 1993; *other-race effect*, Rhodes et al., 2007). This bias in coding means that individuals are tuned by categorization experience to detect subtle differences between other human individuals, thus facilitating face recognition among in-group members at the (individuating) exemplar level (see the *feature-selection hypothesis*, Lewin, 2000). In contrast, individuals code information in the out-group that is more relevant for detection of out-group members, that is, information at the category level. At the category level, the best cognitive processing strategy for discriminating faces would be to code information indicating differences in human likeness along the DHL, thus enhancing discrimination of out-group members (i.e., our avatars). In contrast, a processing strategy that is more suited to face recognition of the individual human category exemplars than processing differences in human likeness along the DHL is more likely to result in poorer discrimination performance for human faces.

Face recognition among in-group members at the individuating level is more likely to rely on the use of configural information (Maurer et al., 2002), whereas there is evidence of less configural coding of out-group members (e.g., Rhodes et al., 1989; Fallshore & Schooler, 1995). Configural information relates to the individual arrangement of first- and second-order (e.g. nose-mouth distance) spatial relations among facial features (Rhodes, 1988). Configural processing is disrupted when faces are inverted instead of being presented upright (Bartlett & Searcy, 1993; Diamond & Carey, 1986; Rhodes et al, 1993; Rossion, 2009). If the asymmetry in PD between the avatar and human faces is attributable to a greater tendency to individuate human category exemplars than avatar category exemplars and a bias therefore toward greater configural processing of human exemplars, then face inversion should reduce or eliminate the asymmetry. If on the other hand the asymmetry is not attributable to differences in configural processing, face inversion will have no impact on it. Experiment 3 was performed in order to test this.

5.5. Experiment 7. Perceptual discrimination and differential processing bias

5.5.1. Materials and Methods

The ABX and forced choice categorization tasks were performed. The task requirements, instructions and stimulus presentation conditions for these tasks were identical to those described for the two preceding experiments, with one exception. The re-morphed stimuli that were presented in Experiment 2 were inverted by rotating them 180°.

Participants

A new sample of $N = 25$ volunteers (21 female, mean age 21 years; range 18 - 26 years) not involved in Experiments 1 or 2 participated in Experiment 3.

5.5.2. Results

Forced choice categorization task: Logistic function, and category boundary

The parameter estimates derived from each logistic function model of each participant across continua were tested against zero in a one-sample t-test and showed, as in Experiments 1 and 2, a highly significant logistic component ($t_{24} = 22.29$, $p > 0.001$) (see Fig. 7). Based on the parameter estimates β_0 and β_1 , the mean category boundary value was $M = 6.7$. Across continua, the data show that the most ambiguous face morph M6 is closest to this boundary (Fig. 7).

For completeness, the other analyses for the forced choice categorization task conducted in Experiments 1 and 2 (i.e., categorization responses and RT) were repeated for Experiment 3. These produced the same pattern of results as Experiments 1 and 2.

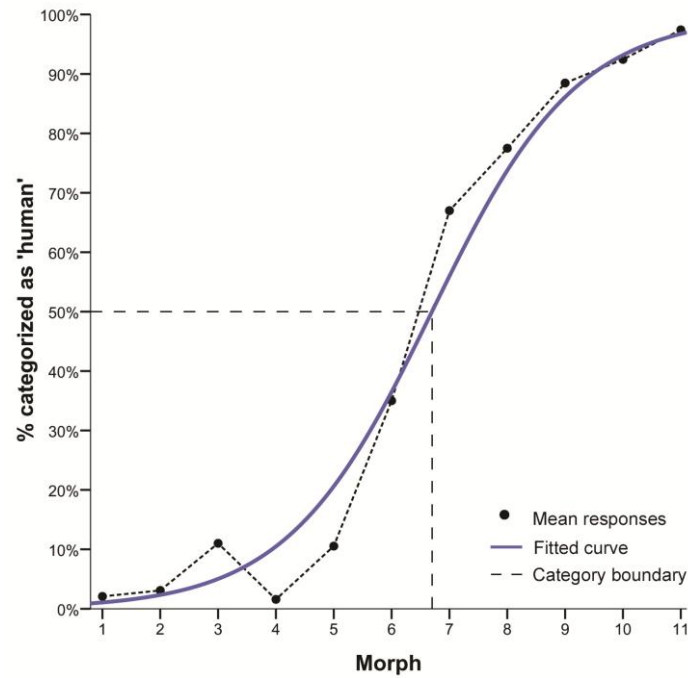


Figure 7. Results of the forced choice classification task for inverted faces. Mean responses are depicted in terms of % of “human” responses (A), with the grand average (continuous blue line) across all continua (with error bars showing standard deviation), the fitted logistic curve based on the grand mean (black line), and the category boundary (dashed gray line) to indicate the point of maximum uncertainty of 50% in categorisation judgements along the continua. Results indicate a step-like response function consistent with the presence of a category boundary.

ABX perceptual discrimination task

An independent two-sample t-test (Experiment 3 versus Experiment 2) using mean d' of each participant across continua as dependent variable was conducted to compare perceptual discrimination performance in Experiments 2 and 3; these were compared because these experiments used the same re-morphed continua. This analysis showed that discrimination performance for each of the 9 morph pairs (i.e., M1-3 to M9-11) was not significantly different between Experiments 2 and 3. Levene's test of equality of variances indicated that the group variances for each of the 9 morph pairs can be treated as equal. For completeness, the same analysis was repeated to test for differences between Experiment 3 versus Experiment 1. This showed a significant difference in discrimination between morph pairs M6-M8 [$t_{72}=2.12$, $p > 0.038$] (note that M7 in Experiment 1 and M6 in Experiment 3 were the most

ambiguous) and between the most human morph pairs M9-M11, $t_{72} = 3.5$, $p > 0.001$. There were no other differences.

PD performance in Experiment 3 was then tested. Given that face morph position M6 was closest to the category boundary, differences in the ability to perceptually discriminate between pairs of morphs (M5-M7) straddling the *ambiguous* M6 was compared with the ability to perceptually discriminate between unambiguous morphs within the *avatar* (M1-M3, M2-M4, M3-M5) and *human* (M7-M9, M8-M10, M9-M11) face categories were tested. This choice of morph pairs was based on the preceding data of the forced choice task, and ensured control for the physical morph distance between the ambiguous and unambiguous faces. The mean value of discrimination sensitivity was compared in a one-way RM-ANOVA with factor morph position (3 levels: avatar, ambiguous, human) using d' as dependent variable. This analysis showed a significant effect for morph pair position, $F(2, 48) = 11.18$, $p < 0.001$.

Tests of planned within-subject contrasts showed the same pattern of significant differences in PD as in Experiment 2. PD of avatar faces ($M = 1.44$, $SE = 0.13$) was significantly greater than that of faces within the human category ($M = 0.87$, $SE = 0.11$), $F(1, 48) = 27.31$, $p > 0.001$. As in Experiment 2, discrimination sensitivity for ambiguous faces at the category boundary ($M = 1.32$, $SE = 0.15$) was not significantly greater than for faces within the avatar category, $F(1, 24) = 1.03$, $p = 0.319$, but it was significantly greater than for faces within the human category, $F(1, 24) = 9.5$, $p = 0.005$.

The data thus indicate that face inversion had no differential impact on the ability to discriminate between faces along the continua.

A one-way RM ANOVA with "morph position" (11 levels) and c as the dependent variable for response bias showed no significant differences for c .

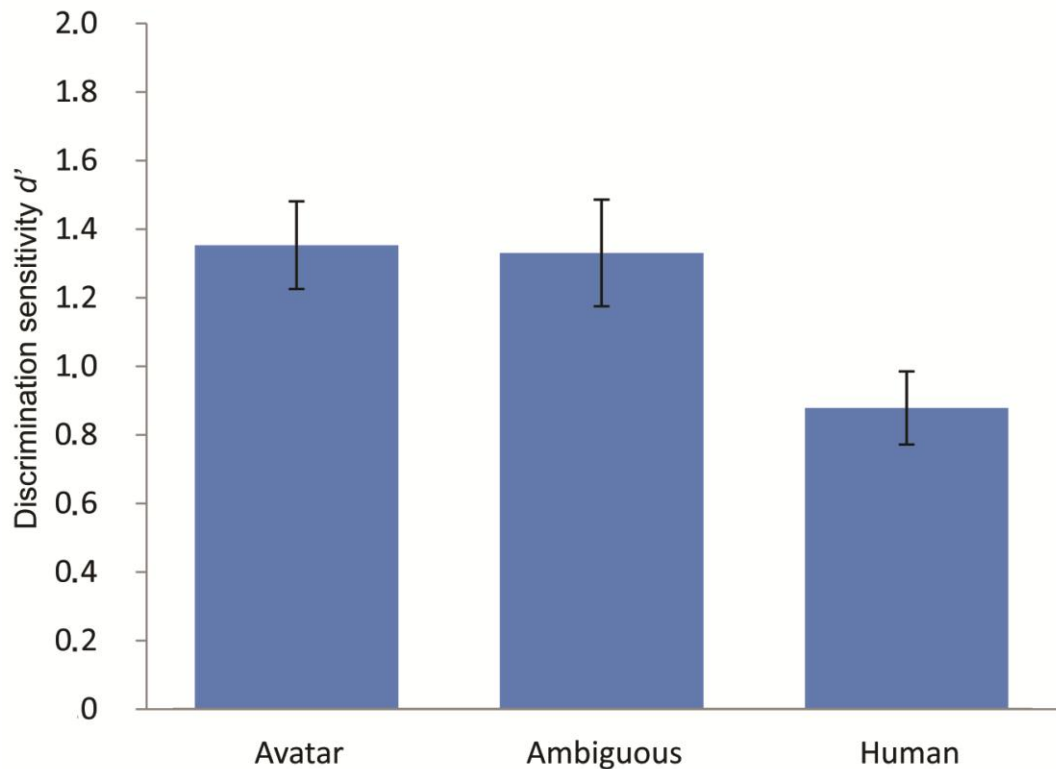


Figure 8. Results of the ABX perceptual discrimination task for inverted faces. This figure depicts that mean discrimination sensitivity d' in the ABX perceptual discrimination task for inverted unambiguous avatar and human and highly ambiguously faces. The data replicate those of experiments 1 and 2, showing the same asymmetry in perceptual discrimination performance along the dimension of human likeness. Face inversion had no impact on this, indicating that this asymmetry is not attributable to a differential processing strategy in which avatars are coded at a category and human faces at an individual level. The error bars indicate 1 SE ($N = 25$).

5.5.3. Discussion

Experiment 3 explored the possibility that the asymmetry in PD reported in Experiments 1 and 2 might be attributable to a *differential processing bias*. This bias suggests that participants preferentially code human-category exemplars at the individual level and avatar-category exemplars at the category level. The data show that the inversion of faces did not impact the asymmetry in PD, indicating that the asymmetry is not likely to be attributable to differences in configural coding and to a tendency to process human compared with avatar faces at an individual level.

5.6. General discussion

The UVH conceptualizes the DHL as a linear dimension of physical similarity space. This space is considered to span between points representing similar objects or characters of various degrees of human likeness within the nonhuman category and a single point representing the human category (Fig.1). The problem with this conceptualization and, more importantly, its faithful application in uncanny studies and theoretical considerations (e.g., Tinwell and Grimshaw, 2009; Ramey, 2005) is that it implicitly assumes that this space does not vary within the human category. The assignment of physically different morphs to the human category in the forced choice categorization task clearly shows that this assumption is wrong (see also e.g., Yamada et al., 2012, Cheetham et al., 2011; Looser and Wheatley, 2011).

The advantage of considering the human end of the DHL is that this provides a basis of comparison for understanding how other objects and characters along the DHL are perceived and experienced. This approach is important for the present study. The UVH predicts enhanced negative affective experience as a function of enhanced PD difficulty and suggests that this effect occurs at the point along the DHL at which categorization ambiguity is greatest. The data of the first and second experiments confirmed that there are differences in PD performance as a function of human likeness. But the pattern of differences in PD is very different than that implicitly assumed in the UVH. Firstly, and as expected on the basis of previous studies of CP, PD of faces at the category boundary is enhanced compared with PD of within-category human faces. Second, PD of within-category avatars is also enhanced compared with PD of within-category human faces. Together, these findings support the suggested asymmetry in PD along the DHL. Importantly, they show, contrary to the UVH, that PD difficulty is actually greatest for human faces.

On the human side of the DHL's category boundary, this finding is reflected in the warped profile of psychological similarity space that is typically described for CP. This profile is characterized by attenuated perceptual discrimination of faces within the human category compared with enhanced discrimination of faces close to and at the category boundary (e.g., Livingstone et al., 1998). In the present study, warping likely reflects the impact of perceptual and category learning processes over a person's

history of everyday social interactive behaviour with other members of the human category: All participants expressly reported no previous experience with our specific avatar parent faces, no previous experience with similarly humanlike faces (and robots), and no knowledge of previous experience with human likeness-related manipulations of perceptual features such as those applied along our morph continua. In contrast, they considered the human parent faces to be of the kind that they might typically encounter in normal everyday situations. The impact of perceptual and category learning processes is that these likely lead to perceptual desensitization to within-category human features that are therefore perceived as more alike or *equivalent*, and to enhanced perceptual sensitivity close to and at the category boundary to those stimulus features that facilitate assignment of category membership in everyday tasks (e.g., human versus nonhuman). These features are therefore perceived as more *distinctive* (e.g., Campbell et al., 1997; Goldstone, 1994; Lawrence, 1949; Gibson, 1998; for an overview of *acquired distinctiveness* and *acquired equivalence*, see, e.g., Goldstone, 2009). In contrast to the warped profile on the human side of the DHL's category boundary, there was no such difference in PD for unambiguous within-category avatar faces compared with the ambiguous faces at or closest to the category boundary. Considered in terms of the CP literature, participants thus appear to be perceptually desensitized to information that would facilitate visual discrimination of within-category human faces, while a corresponding desensitization is not apparent within the nonhuman category.

The present study did not aim to show that PD within the nonhuman category can change with experience, but stimulus exposure and explicit categorization training is known to evoke changes in discrimination sensitivity to a range of stimuli, from simple line drawings of unnatural entities to perceptually complex facial stimuli (e.g., Livingston et al., 1997; Kikutani et al., 2008, 2010; Stevenage, 1998; Goldstone, Steyvers, & Rogosky, 2003; Goldstone, 1996; Levin, 1996, 2000; Schyns and Murphy, 1994; Gibson, 1991; Hall, 1991). If categorization training can modulate PD along the DHL, this might induce effects of *acquired equivalence* (i.e., decreased perceptual sensitivity for within-category faces), *acquired equivalence* (i.e., increased perceptual sensitivity for faces either side of the category boundary) or both, resulting therefore in a different profile of warping along the DHL than shown in the present study. Presumably, categorization training would primarily influence the cognitive

representation of the avatar side of the DHL. Training could be based, for example, on familiarization with avatar faces so that individuals learn to discriminate between these in terms of their unique features (McGugin et al., 2011; Bruyer et al., 2004). Alternatively, the impact of experience might be examined in designers. Animators, video game designers and roboticists concerned about the uncanny effect and the impact of their designs on subjective affect (e.g., Minato et al., 2006; Walters et al., 2008; MacDorman et al., 2009) regularly expose themselves to a range of humanlike faces and actively engage in carefully crafting perceptual features related to human likeness. Differences between novices and experts in processing perceptual information has been reported for other domains of expertise, ranging from the diagnosis of aberrant structures in x-rays to identification of gender in chickens (e.g., Norman et al., 1992; Peron and Allen, 1988; Myles-Worsley et al., 1988; Biederman and Shiffrar, 1987; Burns and Ward, 1978). This has yet to be examined in the present context.

In view of this asymmetry in PD, the third experiment examined whether avatars are preferentially coded at the category level and human faces at the exemplar level. This idea draws on findings relating to the *other-race affect* that show greater accuracy recognizing individual own- compared with other-race faces and show less configural coding of out-group members (e.g., Rhodes et al., 2007; Rhodes et al., 1989). The third experiment used inverted faces because inversion strongly influences efficient configural coding of spatial relations (e.g. nose-mouth distance) among facial features (Leder and Bruce, 2000), while its impact on processing the individual features is generally much weaker (e.g., Murray et al., 2000). The lack of an inversion effect in the present experiment suggests that PD along the DHL generally relies more on coding human likeness-specifying information of facial features such as the eyes, nose, and mouth and other features such as skin tone rather than on coding the spatial relationship among these features, even though coding configural information might enhance the accuracy of coding facial features (Tanaka and Farah, 1993).

The potential role of these features in the reported asymmetry in PD is therefore worth considering in terms of the *avatar-feature hypothesis* (Cheetham et al., 2013). This hypothesis initially related to categorization performance along the DHL. It

suggests that participants preferentially detect perceptual information in nonhuman faces that is diagnostic of the nonhuman category. Assuming that it is cognitively less demanding to detect the presence of this diagnostic information in avatars rather than its absence in human faces, a categorization decision strategy based on “avatar versus not avatar” instead of “avatar versus human” would result in faster categorization decisions for avatars (see also *feature asymmetry*, Triesmman and Gorman, 1988). Consistent with this, the forced choice categorization data of all three experiments show shorter categorization response latencies for avatar compared with human faces, replicating the data of previous studies (Cheetham et al., 2013, 2011; see also Levin, 1996). It is similarly possible that participants preferentially detected or found it easier to detect perceptual information in nonhuman faces that is diagnostic of human likeness along the DHL, thus facilitating the asymmetric effect in PD for these faces. The absence of an inversion effect in the ABX task indicates that this information is not relational (i.e. based on configural coding). Given that inversion effects are weaker for facial features like the eyes, nose and mouth and absent for facial properties like facial color (Leder and Carbon, 2006), it is conceivable that the participants coded and processed perceptual differences along the DHL on the basis of facial properties such as smoothed skin texture, color and shading. This does not exclude a role for feature-based processing, especially as processing for example the general luminance properties of faces can enhance processing of facial features (Sergent, 1986; Schyns and Oliva, 1994; Schyns and Gosselin, 2003). The question is why these properties should be easier to detect in the avatar faces. In view of the task context of processing novel avatars and everyday human faces, it is possible that perceptual information indicating the novelty of these facial properties renders these more salient in the nonhuman regions of the DHL and that novelty therefore serves as a primitive perceptual feature that can facilitate PD within the avatar category (Levin, 2000).

An alternative explanation is that visual discrimination performance might be facilitated by the progressive reduction in perceptual complexity of the morphs with increasing distance from the human end of the continua independently of experience and perceptual strategy. The avatar parent faces have less structural and textural detail than the human parent faces and it is possible that certain perceptual properties of the avatar morphs such as the reduced variance in shading of the

smoothed skin texture is in itself a readily detectable feature of these morphs that eases discrimination. On the other hand, if enhanced perceptual simplicity toward the avatar end of the continua does drive discrimination one might expect a more linear increase in discrimination performance as a function of morph distance from the human end of the continua and no indication of warping between the category boundary and the human side of the DHL.

The UVH predicts that greater PD will evoke greater feelings of strangeness (i.e., feelings of less familiarity) at the point along the DHL at or near which ambiguity is greatest. The data of the second experiment suggest that this is wrong on two counts. Firstly, the analysis of familiarity ratings indicates that greater feelings of strangeness (i.e., feelings of less familiarity) are not reported for ambiguous faces. Instead, feelings of strangeness increased with increasing morph distance from the human end of the continua. This is consistent with the pattern reported in other studies in which comparably well-controlled morph continua and *ad hoc* measures of shinwakan such as measures of pleasantness have been used (e.g., Looser and Wheately, 2011). This is also consistent with a recent study in which well-validated psychophysiological and self-report measures were used to investigate the impact of category ambiguity on valence and arousal (Cheetham et al., submitted to Frontiers). Valence and arousal were used as these form the primary orthogonal dimensions of affective experience (Schlossberg, 1954; Russell, 1980; Russell, 2003; Yik et al., 1999; Posner et al., 2005). Based on these dimensions, Cheetham et al.'s study shows no evidence to suggest any uncanny-like effect along the DHL. That study also applied the same familiarity scale as used in the present study, showing the identical profile in familiarity ratings. This finding is already discussed in detail in that study. But the overall implication of the present familiarity data is that more humanlike stimuli simply evoke more positive affective experience and are preferred over less humanlike stimuli. The most straightforward explanation for this relates to the *mere-exposure* effect (Zajonc, 1968). This means that repeated exposure to human faces over a person's history of social interaction and the often more positive affective tone of interaction with particular in-groups results in more positive evaluations of other in-group members (e.g., Reis and Gable, 2003).

Second, the inter-individual differences approach adopted in the second experiment shows that there is indeed a significant relationship between familiarity and PD difficulty, but that the direction of this relationship is the opposite of that predicted in the UVH. Increasing PD difficulty is associated with more positive feelings of familiarity. The effect was evident for nonhuman and ambiguous faces, whereas there was no significant relationship between PD and feelings of familiarity for human faces. This correlative effect was greatest for ambiguous faces, indicating that, irrespective of the causal relationship between PD and feelings of familiarity, the UVH' prediction is most likely to be incorrect. It should be noted that the UVH does not suggest how DP difficulty and its affective dimension, *shinwakan*, should be operationalised. This issues has hampered uncanny-related research from the outset. But the approach taken in the present study to testing the relationship between DP and affective experience (as described in the UVH) was straightforward and produces strong effects, indicating that further examination of this relationship might be fruitful.

Why greater PD difficulty should correlate with more positive self assessment of affect is not clear. A popular account of the uncanny effect is based on the *Hedonic Fluency Model* (Winkielman et al., 2003; see Yamada et al., 2012). This suggests that negative evaluations of novel or unfamiliar stimuli relate to cognitive difficulty extracting information needed for rapid and efficient processing. This makes sense if the UVH' prediction for PD is assumed to be correct. But the present data suggest that this prediction is incorrect. The present PD data do however fit better with an alternative model of processing fluency, the *Fluency Amplification Model* (Albrecht and Carbon, 2014). This model states that processing fluency enhances the affective reaction that the stimulus already evokes. Assuming that the valence of a given stimulus is initially experienced as comparatively negative (in our case, the more negatively rated unambiguous avatar and ambiguous faces), individuals who experience greater fluency (in our case, lesser difficulty in PD) will experience the negative stimulus as even more negative. By the same token, greater PD difficulty would correlate with less negative ratings. This is consistent with the effect reported in the present study. This might be an interesting avenue of further investigation, especially as the Fluency Amplification Model is also consistent with the data reported in Cheetham et al. (submitted to Frontiers).

The ABX perceptual discrimination task is useful for testing naive participants because it requires no description of the specific physical dimensions along which the stimuli vary and participants do not need to know the category labels. One explanation for CP effects suggests a role for the presence of category labels (Pilling, Wiggett, Özgen, & Davies, 2003; Roberson & Davidoff, 2000; Kikutani et al, 2008). This is because within-category stimuli differ only at the exemplar level, while cross-category stimuli differ both at the exemplar and category levels. If exemplar-level information and category-level information are processed in parallel so that the category boundary can be represented in naive participants after initial learning (Marsolek, 2004), category-level processing might encourage the use of labeling and the emergence of CP effects. This effect might be even stronger in a task with a strong memory component such as the ABX task (i.e., the test stimulus must be compared with the stored representation of the target stimuli). Considering the asymmetry in discrimination performance around the category boundary, a labeling effect is unlikely, unless labeling affected the human side of the category boundary only. It has however been argued that any impact of category labeling would be reflected in specific *within-category discrimination asymmetries* (Hanley and Roberson, 2011). There are no such asymmetries within the human or nonhuman categories.

In summary, the data of the three experiments reject the implicit assumption underlying the UVH' key prediction. The data show lesser PD difficulty for categorically ambiguous faces and for unambiguous avatar faces and, notably, greater PD difficulty for unambiguous human faces. The data indicate that this asymmetry in PD difficulty cannot be attributed to differences between human and nonhuman faces in configural coding. It is likely that perceptual differences along the DHL are generally processed on the basis of human likeness-related manipulations of facial properties such as skin texture, color and shading. Ratings of familiarity show that faces associated with greatest category ambiguity do not show an uncanny-like effect. Negatively valenced ratings increased across the tested stimulus conditions with increasing distance from the human end of the continua. An interindividual differences approach revealed that greater PD difficulty is associated with more positively rather than negatively valenced experience. This challenges the

key idea behind the UVH. This effect is strongest for ambiguous faces, suggesting that this effect is more consistent with the metaphor “happy valley”. These findings for our static faces thus indicate that both the assumed distribution of PD difficulty along the DHL and the predicted relationship between PD difficulty and affective experience (as ‘conceptualized’ in the UVH) are wrong.

Clearly, it is not possible to refute the vaguely formulated non-scientific UVH in its current form. Our approach has been to augment the notions underlying the UVH with the necessary assumptions needed to render the essential features of the hypothesis testable. While we find no evidence in favor of these notions, our findings do not exclude the possibility that other well-controlled experimental paradigms and other methods show effects consistent with the prediction of the UVH. It should be noted that only male face stimuli were presented. The choice of stimuli for this study was not guided by the well-known depiction of Mori’s hypothesis in Figure 1, because we sought to ensure that perceptual discriminative, categorization and familiarity judgments would not be confounding by factors other than degrees of human likeness in the depiction (for a discussion of confounds, see Cheetham and Jancke, 2013). This study presented stimuli similar to those used in preceding studies (Cheetham et al., 2011, 2013). Given the absence of comparable investigations of the DHL, this approach has provided an effective means to developing a basis for further uncanny-related investigation. But an important element of further such study would be to examine whether these findings generalize to other stimuli.

Conflict of Interest Statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgments

This work was supported by the European Union FET Integrated Project PRESENCCIA (Contract number 27731).

5.7. References

- Annett, M. A classification of hand preference by association analysis. (1970). *Br J Psychol.*, 61(3):303-21
- Bartlett, J. C., & Searcy, J. (1993). Inversion and configuration of faces. *Cognitive Psychology*, 25, 281-316.
- Brehaut, J. C., & Tipper, S. P. (1996).
Beale J.M., & Keil F.C. (1995). Categorical effects in the perception of faces. *Cognition*, 57, 217–219.
- Best, C.T., Morrongiello, B., & Robson, R. (1981). Perceptual equivalence of acoustic cues in speech and nonspeech perception. *Perc. & Psychophys.*, 29, 191–211.
- Bruyer, R., Leclere, S., & Quinet, P. (2004). Ethnic categorisation of faces is not independent of face identity. *Perception*, 33(2): 169–179. doi: 10.1068/p5094.
- Bülthoff, I., & Newell, F.N. (2000): Investigating categorical perception of gender with 3-D morphs of familiar faces. *Perception*, 29, 57
- Burleigh, T. J., Schoenherr, J. R., & Lacroix, G. L. (2013). Does the uncanny valley exist? An empirical test of the relationship between eeriness and the human likeness of digitally created faces. *Computers in Human Behavior*.
<http://dx.doi.org/10.1016/j.chb.2012.11.021>.
- Campbell, R., Pascalis, O., Coleman, M., Wallace, S.B., & Benson, P.J. (1997). Are faces of different species perceived categorically by human observers? *Proc. R. Soc. Lond. B: Biol. Sci.*, 264, 1429–1434.
- Cheetham, M., & Jancke, L. (2013). Perceptual and category processing of the 'Uncanny Valley Hypothesis' dimension of human likeness: some methodological issues. *J. Vis. Exp.* (76), e4375, doi:10.3791/4375
- Cheetham, M., Suter, P., & Jancke, L. (2011). The human likeness dimension of the "uncanny valley hypothesis": behavioral and functional MRI findings. *Front Hum Neurosci.*, 5, 126.
- Cheetham, M., Pavlovic, I., Jordan, N., Suter, P., & Jancke, L. (2013). Category Processing and the human likeness dimension of the Uncanny Valley Hypothesis: Eye-Tracking Data. *Front Psychol.*, 4, 108. doi: 10.3389/fpsyg.2013.00108.
- Corneille, O., & Judd, C. M. (1999). Accentuation and sensitization effects in the categorization of multi-faceted stimuli. *Journal of Personality and Social Psychology*, 77, 927-941.
- Corneille, O., Goldstone, R.L., Queller, S., & Potter, T. (2006). Asymmetries in categorization, perceptual discrimination, and visual search for reference and nonreference exemplars. *Mem Cognit.*, 34(3), 556-567.
- Diamond, R., & Carey, S. (1977). Development changes in the representation of faces. *Journal of Experimental Child Psychology*, 23, 1-22.
- Diamond, R., & Carey, S. (1986). Why faces are and are not special: an effect of expertise. *Journal of Experimental Psychology: General*, 115(2), 107–117.
- Ectoff, N.L., & Magee, J.J. (1992). Categorical perception of facial expressions. *Cognition*, 44, 227-240.
- Fabri, M., Moore, D.J., & Hobbs, D.J (2004) Mediating the Expression of Emotion in Educational Collaborative Virtual Environments: An Experimental Study. *International Journal of Virtual Reality*, Springer Verlag, London
- Fallshore, M., & Schooler, J. W. (1995). The verbal vulnerability of perceptual expertise. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 21, 1608-1623.
- Francis, A.L., & Ciocca, V. (2003). Stimulus presentation order and the perception of lexical tones in Cantonese. *JASA.*, 114, 1611–1621.

- Gelder, B. de, Teunisse, J., P., & Benson, P., J. (1997). Categorical perception of facial expressions: Categories and their internal structure. *Cognition and Emotion*, 11(1) 1 - 23.
- Gibson, E. J. (1991). *An odyssey in learning and perception*. MIT Press: Cambridge.
- Goldstone, R. (1994). Influences of categorization of perceptual discrimination. *Journal of Experimental Psychology: General*, 123, 178-200.
- Goldstone, R. L. (1998). Perceptual learning. *Annual Review of Psychology*, 49, 585–612.
- Goldstone, R. L., Lippa, Y., and Shiffrin, R. M. (2001). Altering object representations through category learning. *Cognition*, 78(1), 27-43.
- Gosselin, F., Archambault, A., & Schyns, P.G (2001) Interactions between taxonomic knowledge, categorization, and perception. In: Hahn, U. and Rascar, M. (eds.) *Similarity and Categorization*. Oxford University Press.
- Hall, G. (1991). *Perceptual and associative learning*. Clarendon Press: Oxford.
- Hanley, J.R. & Roberson, D. (2011) Categorical perception effects reflect differences in typicality on within-category trials. *Psychon Bull Rev.*, 18(2), 355-63. doi: 10.3758/s13423-010-0043-z.
- Hanson, D. (2006) “Exploring the aesthetic range for humanoid robots”, in *Proceedings of the ICCS/CogSci-2006 Long Symposium: Toward Social Mechanisms of Android Science*, pp. 39–42.
- Harnad, S. (1987). Categorical perception. Cambridge University Press: Cambridge.
- Hautus, M. J. & X. Meng (2001). Decision strategies in the ABX (matching-to-sample) psychophysical task. *Perception & Psychophysics*, 64, 89-106.
- Ho, C.-C., & MacDorman, K. (2010). Revisiting the uncanny valley theory: developing and validating an alternative to the Godspeed indices. *Comput. Hum. Behav.* 26, 1508–1518.
- Johansen MK, Palmeri TJ. (2002). Are there representational shifts during category learning? *Cogn Psychol.*, 45(4), 482-553.
- Kikutani, M., Roberson, D., & Hanley, J. R. (2008). What’s in the name? Categorical perception of unfamiliar faces can occur through labeling. *Psychonomic Bulletin & Review*, 15, 787–794.
- Leder, H. & Bruce, V. (2000). When inverted faces are recognized: The role of configural information in face recognition. *Quarterly Journal of Experimental Psychology Section A-Human Experimental Psychology*, 53, 513-536.
- Leder, H. & Carbon, C. C. (2006). Face-specific configural processing of relational information. *British Journal of Psychology*, 97, 19-29.
- Levin, D. T. (1996). Classifying faces by race: The structure of face categories. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 22, 1364-1382.
- Levin, D. T. (2000). Race as a visual feature: Using visual search and perceptual discrimination tasks to understand face categories and the cross race recognition deficit. *Journal of Experimental Psychology: General*, 129, 559-574.
- Levin, D. T., & Beale, J. M. (2000). Categorical perception occurs in newly learned faces, other-race faces, and inverted faces. *Perception & Psychophysics*, 62, 386-401.
- Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, 54, 358–368.
- Livingston, K. R., Andrews, J. K., & Harnad, S. (1998). Categorical perception effects induced by category learning. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 24(3), 732– 753.

- Looser, C.E., & Wheatley, T. (2010). The tipping point of animacy: How, when, and where we perceive life in a face. *Psychological Science*, 21, 1854–1862
- MacDorman, K. (2005). "Androids as an Experimental Apparatus: Why is There an Uncanny Valley and Can We Exploit it?" in *Toward Social Mechanisms of Android Science: A CogSci 2005 Workshop*, Stresa, 106–118
- MacDorman, K. F. & Ishiguro, H. (2006). The uncanny advantage of using androids in cognitive science research. *Interaction Studies*, 7(3), 297-337.
- MacDorman, K., Green, R., Ho, C.-C., & Koch, C. (2009). Too real for comfort? Uncanny responses to computer generated faces. *Computers in Human Behavior*, 25(3), 695– 710
- Macmillan, N. A., & Creelman, C. D. (2005). *Detection Theory: A User's Guide*. Mahwah, New Jersey : Lawrence Erlbaum Associates .
- Marsolek, C. (2004). Abstractionist versus exemplar-based theories of visual word priming: A subsystems resolution. *Quarterly Journal of Experimental Psychology*, 57A, 1233-1259.
- Massaro, D.W. & Cohen, M.M. (1983). Categorical or continuous speech perception: A new test. *Speech Communication*, 2, 15-35.
- Matsuda, Y.T., Okamoto, Y., Ida, M., Okanoya, K., & Myowa-Yamakoshi, M. (2012). Infants prefer the faces of strangers or mothers to morphed faces: an uncanny valley between social novelty and familiarity. *Biol Lett.*, 8(5), 725-8. doi: 10.1098/rsbl.2012.0346.
- Maurer, D., Le Grand, R., & Mondloch, C.J. (2002) The Many Faces of Configural Processing. *Trends in Cognitive Sciences* 6: 255–260. doi:10.1016/s1364-6613(02)01903-4.
- McGugin, R.W., Tanaka, J.W., Lebrecht, S., Tarr, M.J., & Gauthier, I. (2011). Race-specific perceptual discrimination improvement following short individuation training with faces. *Cogn Sci.*, 35(2), 330-347. doi: 10.1111/j.1551-6709.2010.01148.x.
- Minato, T., Shimada, M., Itakura, S., Lee, K., & Ishiguro, H. (2006). Evaluating human likeness of an android by comparing gaze behaviors elicited. *Advanced Robotics*, 20, 1147-1163.
- Moore, R.K. (2012). A Bayesian explanation of the 'Uncanny Valley' effect and related psychological phenomena. *Scientific Reports* 2, 864. doi:10.1038/srep00864
- Mori, M. 1970. Bukimi no tani [The uncanny valley]. *Energy*, 7(4) 33-35. (Translated by Karl F. MacDorman and Takashi Minato in 2005 within Appendix B for the paper *Androids as an Experimental Apparatus: Why is there an uncanny and can we exploit it?* In *Proceedings of the CogSci-2005 Workshop: Toward Social Mechanisms of Android Science* (Italy, July 2005), 106-118).
- Murray, J. E., Yong, E., & Rhodes, G. (2000). Revisiting the perception of upside-down faces. *Psychological Science*, 11(6), 498–502.
- Myles-Worsley, M., Johnston, W.A., & Simons, M.A. (1988). The influence of expertise on X-ray image processing. *J. Exp. Psychol.: Learn. Mem. Cogn.*, 14, 553-557
- Newell, F.,N., & Bulthoff, H.H. (2002). Cognition Categorical perception of familiar objects. *Cognition*, 85, 113–143.
- Notman, L.A., Sowden, P.T., & Ozgen, E. (2005). The nature of learned categorical perception effects: a psychophysical approach. *Cognition*. 95(2), 1-14.
- Patterson, C.H. (1983). *Theories of counseling and psychotherapy*. Philadelphia: Harper & Row.
- Popper, K. (1992). *The Logic of Scientific Discovery*. London: Routledge.

- Popper, K. (2004). *Conjectures and refutations: the growth of scientific knowledge*. London: Routledge.
- Pierce, J. R., & Gilbert, E. N. (1958). On AX and ABX limens. *Journal of the Acoustical Society of America*, 30, 593-595.
- Pilling, M., Wiggett, A., Özgen, E., & Davies, I. R. L. (2003). Is color "categorical perception" really perceptual? *Memory & Cognition*, 31, 538-551.
- Pisoni, D. B., (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics*, 13, 253-260.
- Pollick, F. E. (2006). In search of the uncanny valley. In K. Grammer, & A. Juett (Eds.), *Analog communication: Evolution, brain mechanisms, dynamics, simulation*. Cambridge, MA: MIT Press, The Vienna Series in Theoretical Biology.
- Prins, N., & Kingdom, F. A. A. (2009). *Palamedes: Matlab routines for analyzing psychophysical data*. <http://www.palamedestoolbox.org>
- Ramey, C. H. (2005). The uncanny valley of similarities concerning abortion, baldness, heaps of sand, and humanlike robots. Paper Presented at the Proceedings of *Views of the Uncanny Valley Workshop: IEEE-RAS International Conference on Humanoid Robots*, Tsukuba.
- Reis, H. T., & Gable, S. L. (2003). Toward a positive psychology of relationships. In C. L. Keyes & J. Haidt (Eds.), *Flourishing: The positive person and the good life* (pp. 129–159). Washington, DC: American Psychological Association
- Repp, B. H. (1984). Categorical perception: Issues, methods and findings. In N. Lass (Ed.), *Speech and language, vol. 10: Advances in basic research and practice* (pp. 244–335). Orlando, FL: Academic Press
- Rhodes, G. (1988). Looking at faces: First-order and second-order features as determinants of facial appearance. *Perception*, 17(1), 43–63.
- Rhodes, G., Brake, S., & Atkinson, A. P. (1993). What's lost in inverted faces? *Cognition*, 47(1), 25-57.
- Rhodes, G., Tan, S., Brake, S., & Taylor, K. (1989). Expertise and configural coding in face recognition. *British Journal of Psychology*, 80, 313–331.
- Roberson, D., & Davidoff, J. (2000). The categorical perception of colors and facial expressions: The effect of verbal interference. *Memory & Cognition*, 28, 977-986.
- Rossion, B. (2009). Distinguishing the cause and the consequence of face inversion: The perceptual field hypothesis. *Acta Psychologica*, 132, 300–312.
- Schyns, P. G. & Murphy, G. L. (1994). The ontogeny of part representation in object concepts. In Medin (Ed.), *The Psychology of Learning and Motivation*, 31, 301-349.
- Schyns, P. G. (1991). A modular neural network of concept acquisition. *Cognitive Science*, 15, 461-508
- Schyns, P.G., & Oliva, A. (1994). From blobs to boundary edges: evidence for time and spatial scale dependent scene recognition. *Psychol. Sci.* 5, 195–200.
- Schyns, P.G., & Gosselin, F. (2003). "Diagnostic use of scale information for componential and holistic recognition," in *Perception of Faces, Objects, and Scenes. Analytic and Holistic Processes*, eds M. A. Peterson and G. Rhodes (Oxford: Oxford University Press), 120–145.
- Sergent, J. (1986). "Microgenesis of face perception," in *Aspects of Face Processing*, eds H.D. Ellis, M.A. Jeeves, F. Newcombe, and A.M. Young (Dordrecht: Martinus Nijhoff), 17–33.

- Sigala, R., Logothetis, N.K., & Rainer, G. (2011). Own-species bias in the representations of monkey and human face categories in the primate temporal lobe. *Journal of Neurophysiology*, 105(6), 2740-2752.
- Stevenage, S. V. (1998). Which twin are you? A demonstration of induced categorical perception of identical twin faces. *British Journal of Psychology*, 89, 39-57.
- Studdert-Kennedy, M., Liberman, A.M., Harris, K.S., & Cooper, F.S. (1970). Motor theory of speech perception: A reply to Lane's critical review. *Psychological Review*, 77(3), 234-249.
- Tanaka, J. W., & Curran, T. (2001). A neural basis for expert object recognition. *Psychological Science*, 12, 43-47.
- Tinwell A, Grimshaw, M, & Williams A, (2010). Uncanny behaviour in survival horror games. *Journal of Gaming & Virtual Worlds* , 2(1), 3-25.
- Tinwell, A. 2009. The Uncanny as Usability Obstacle. In Proceedings of the “Online Communities and Social Computing” workshop, *HCI International 2009*. Springer, Volume 12, San Diego, CA.
- Tinwell, A., & Grimshaw, M. (2009). Bridging the uncanny: an impossible traverse? In O. Sotamaa, & A. Lugmayr, H. Franssila, P. Näränen, & J. Vanhala (Eds.), *Proceedings of the 13th International MindTrek Conference: Everyday Life in the Ubiquitous Era* (pp. 66-73). Tampere, Finland: ACM.
- Treisman, A, & Gormican S. (1988). Feature analysis in early vision: evidence from search asymmetries. *Psychol Rev.* 95(1), 15-48.
- Valentine, T., & Endo, M. (1992). Towards an exemplar model of face processing: The effects of race and distinctiveness. *Quarterly Journal of Experimental Psychology*, 44A, 671-703.
- Van Selst, M. & Jolicoeur, P. (1994). A solution to the effect of sample size on outlier elimination. *Quarterly Journal of Experimental Psychology*, 47A, 631-650.
- Viviani, P., Binda P., & Borsato T. (2007). Categorical perception of newly learned faces. *Visual Cognition* , 15, 420–467

6. General discussion

On the basis of informal observation of individuals who showed signs of personal disquiet when presented with humanlike robots, Mori (1970) speculated about a phenomenon that might pose a general problem for designers of very realistic anthropomorphic characters and objects (e.g. robots, mannequins, or prostheses). This problem is expressed in the non-scientific statement referred to as the UVH (Popper, 1963; for bold conjecture, see Popper, 1992, 2004). Formulated as a question, the UVH asks what impact highly realistic non-human objects designed to closely simulate human appearance and behaviour have on the affective experience and behaviour of the human observer. The UVH presents this question in the form of a (visual) hyperbole (McQuarrie and Mick, 1996). This uses the exaggerated valley in the shape of the curve and the illustrative description of this to convey a general sense of the problem. While the seeming parsimony of the UVH might contribute to

its persuasiveness on thinking guided by the UVH (for parsimony in well-developed theories, see e.g. Patterson, 1986). But the UCH does not specify any details, embed the proposed problem in contemporary theories and knowledge, or consider the potential complexity of the psychological processes underlying experience of the proposed uncanny effect. As such the UVH represents an early stage of scientific enquiry (Hempel, 1945; Northrop, 1966; Poundstone, 1988). It is in this context that the studies of the present dissertation sought to gain insight into the meaning and relevance of the UVH.

The main prediction of the UVH is that positive or negative affect will be evoked by nonhuman entities as a function of their perceived human likeness. The way in which human likeness is perceived along the DHL is thus critical to understanding and examining the UVH. A valuable approach to examination is to use morph continua to represent the DHL (e.g. MacDorman and Ishiguro, 2006; Hanson et al., 2005; MacDorman et al., 2008; Seyama and Nagayama, 2007; Hanson, 2006). These allow the relationship between fine-grained differences in humanlike appearance and behavioural measures of subjective perception and experience (e.g. category decisions, uncanny feelings) and with underlying neural processes to be examined (Cheetham et al., 2011).

The use of morph continua to represent the DHL in a forced choice categorisation task enabled experiment 1 to demonstrate that *physical similarity space* along the DHL includes variation in human-like appearance within the human category. The UVH and its depiction focuses on physical human-like variation of nonhuman objects and treats human image as a general point of reference (see also Minato et al., 2006; Ramey (2005); Walters et al., 2008). The relevance of this finding is that it affirms the importance of considering differences in human likeness within the human category and its cognitive representation. For the present study, this is important because the UVH prediction deals with the ambiguities in *perceptual discrimination (PD)* or categorisation between the nonhuman and human categories that need therefore to be defined. Experiment 1 defined the point of greatest ambiguity as the category boundary. One approach to investigating the UVH is to augment the vaguely formulated UVH by proposing that any personal discomfort associated with processing human likeness is most likely to occur in response to stimuli at or near the

category boundary at which category decision ambiguity is greatest. Yamada et al., (2012) tested this proposal, finding evidence in support of it.

An alternative reading of the vaguely formulated UVH is that enhanced difficulty perceptually discriminating between human and humanlike characters will evoke negatively valenced experience. Experiment 2 was used in conjunction with the data of experiment 1 to test for effects of categorical perception (CP; see Harnad, 1987). Using a variant of the same-different discrimination task (e.g., Angelini et al., 2008), experiment 2 showed enhanced discriminative sensitivity for stimuli that straddle the category boundary. This finding is relevant for the UVH in two ways. Firstly, it shows that the definition of the DHL in terms of physical similarity space diverges from its cognitive representation in psychological similarity space. This divergence means that after controlling for physical differences, stimuli within certain regions of the DHL (i.e., the human and nonhuman categories) are perceived as more similar and stimuli straddling the category boundary as less similar. This is consistent with the idea that the cognitive representation of *psychological similarity space* (such as along a perceptual dimension like the DHL) can be selectively deformed (Livingstone et al., 1998) and suggests that this effect applies also to the DHL. This is not unique for the DHL.

Second, the UVH proposes that difficulty perceptually discriminating between such stimuli evokes negative affect. The present finding shows however that perceptually discrimination difficulty is in fact greatest for within category stimuli that, according to the UVH, should actually evoke positive affect. This suggests that the UVH might be incorrect. Many studies of CP such as for famous faces (Beale and Keil, 1995), unfamiliar faces (Levin and Beale 2000), facial expressions (Ectoff and Magee 1992) and faces of different gender (Bülthoff and Newell, 2000) show such effects. In terms of the criteria for CP (see Studdert-Kennedy et al., 1970), Experiment 1 confirms that the first criterion for the presence of CP is fulfilled, namely that there is a category boundary. Experiment 2 confirms better discrimination accuracy for pairs that cross the category boundary than for equidistant pairs drawn from within a category, thus indicating a discrimination boundary, that is, the second criterion for CP. It might be assumed that the third criterion is fulfilled, namely that the discrimination boundary is aligned with the category boundary. Critically, the perceptual discrimination task was

based on the use of a four-step procedure. This means that the task required perceptual discrimination between morphs that were four steps apart along the continuum (i.e. the morphs 0, 4, 8, and 12 of the 13-morph continua). This in turn means that the precise profile of perceptual discrimination performance is not delineated. This limitation applies also to the examination of the DHL in Looser and Wheatly (2011) and for studies of CP effect in other dimensions of human likeness (based on human, cow and monkey faces) (Campbell, et al, 1997). The fourth criterion is that discrimination is at chance within the categories (note that this is not always applied, Harnad, 1987; Repp, 1984). Discrimination in this experiment was slightly above chance within the categories (a finer-grained delineation of perceptual discrimination along the DHL was conducted in Experiments 5 to 7).

The use of the four-step procedure in Experiment 2 served a particular purpose. Mori described perceptual uncertainty (and associated uncanny experience) as occurring at levels of realism that correspond to the region along the DHL between the two positive peaks in the slope of the familiarity-human likeness relationship (see Fig. 1). At these peaks, objects are regarded as either nonhuman or human. In reframing Mori's considerations in terms of category processing, these peaks may be seen as reflecting degrees of human likeness at which correctly classified category instances (i.e. nonhuman and human) straddle the category boundary. The UVH does not specify how efficient classification at these peaks must be, though the identification of objects at each peak is clearly considered to be relatively efficient and effortless. Morphs 4 and 8 were tested in Experiment 1 and verified as clearly identifiable as nonhuman and human, respectively (for less conservative criterion indicating each category see e.g. Beale and Keil, 1995; Ectoff and Magee, 1992). Morph M4 was identified on average as an avatar in more than 85% of trials and morph M8 as a human in more than 85% of trials. Please note that this criterion had to apply to both morphs of any one continuum. Using this approach, this choice of morphs sought to capture a sense of category change along the DHL between nonhuman and human objects in accordance with Mori's description.

Experiment 3 used fMRI methodology and the same stimuli and stimulus presentation paradigm (pair-repetition paradigm) as used in Experiment 2 to examine effects of category processing on neural activity, while simultaneously controlling for

effects of physical change along the DHL. The aim was to recreate the stimulus conditions described by Mori in the UVH. Mori suggests that viewing a nonhuman stimulus that is initially perceived to be human evokes negative affect. To this end The fMRI study utilized a phenomenon called *repetition suppression (RS)* (e.g., Grill-Spector et al., 2006; Jiang et al., 2006, see also Saygin et al, 2011, who used the same procedure to examine perception of robots). Pairs of morphs were presented in rapid succession. The repetition in the second morph of stimulus attributes already presented in the first morph of a given pair leads to attenuation in activation (i.e., RS) in those brain region's that are sensitive to that particular attribute. This attenuation can be used to identify the sensitive brain regions. By experimentally manipulating the category of each morph of a morph pair, the effect on neural processing of viewing two faces drawn from within a category can be compared with the effect for faces of a pair that change category. The data showed that a change in category between morph in the avatar-to-human direction along the DHL evoked a different pattern of neural activity (i.e. hippocampus, amygdala, and insula) than a change in category in the human-to-avatar direction (i.e., putamen, head of caudate, thalamus, and red nucleus). At a broad level, this pattern of differences indicates that processing category change along the DHL is differentially impacted by the category of the priming (i.e., first morph of a pair). This suggests that novel avatar and everyday human faces represent different categorization problems for which the process of categorisation relies on the coding of different perceptual features.

Experiment 4 considered this notion and showed that categorically unambiguous avatars are processed differently than ambiguous or human faces as indicated by the relative amount of dwell time spent extracting visual information from the eyes, nose, and mouth. The behavioural data support this by showing that category assignment of avatar faces is also much faster in comparison with the assignment of human faces. Experiment 5 followed the conclusions of experiments 1-4, assuming greater differences perceptual discrimination performance between the nonhuman and human categories than Experiment 2 suggested. Experiment 5 (and 6 and 7) used a perceptual discrimination task (i.e. the ABX task) in which the morph distance between pairs of morphs was reduced to create a two-step procedure (Liberman et al, 1957). The aim of this was to establish clearly whether and where along the DHL differences in PD difficulty as a function of human likeness occur. The reason for this

was a further reading of the vague UVH. This reading predicts that that perceptual difficulty discriminating between highly realistic humanlike objects and characters and their human equivalent will evoke an unpleasant affective state. Critically, the degree of high realism at which this state should occur is conjectured to be at the point along the DHL at which attribution of objects and characters to the human or nonhuman category is subject to greatest ambiguity (i.e. at or near the category boundary).

Experiments 4 and 5 confirmed that there are differences in PD difficulty as a function of human likeness along the DHL, but that the pattern of PD is different than indicated in the UVH. PD difficulty was lowest for characters at or near the category boundary and greatest for human faces. Experiment 5 showed that overall, negative affect increases with increasing distance from the human end of the DHL. Using an inter-individual differences approach to examine the relationship between familiarity and PD difficulty, experiment 5 showed also that greater PD difficulty is associated with enhanced positive affect (as measured in terms of the UVH' dimension of familiarity). This effect was more consistent with the Fluency Amplification Model (Albrecht and Carbon, 2014) rather than the Hedonic Fluency Model of perceptual fluency that has been used to account for uncanny-like effects (Yamada et al., 2012). Interestingly, this effect only applies for nonhuman and ambiguous faces. There was no significant relationship between PD difficulty and familiarity for human faces. Critically, this correlative effect was greatest for ambiguous faces.

Conclusion

In conclusion, the experiments using well-controlled morph continua and different methodological approaches show that many aspects of the UVH are in need of revision: 1) The UVH's conceptualization of linear physical similarity space along the DHL diverges strongly from the nonlinear cognitive representation of the DHL, 2) human category exemplars vary in human-like appearance along the DHL, 3) the perceived direction of category change along the DHL evokes different processes or processing strategies during categorization of faces, 4) the underpinning pattern of brain regions suggest that these different processes relate to differences in perceptual and categorization expertise, 5) the potential impact of this is indicated in

differences in PD difficulty as a function of human likeness, 6) critically, this pattern is different than that implicitly assumed in the UVH, with greatest PD difficulty for human faces and least PD difficulty for ambiguous faces, 7) there is no general increase in negative affect (as assessed in terms of the UVH' dimension of familiarity) for categorically ambiguous stimuli along the DHL, 8) negative affect increases with increasing distance from the human end of the DHL, 9) greater PD difficulty is associated with enhanced positive affect (i.e., greater feelings of familiarity) instead of less positive affect (i.e., greater feelings of strangeness) as the UVH suggests, and, 10) presenting a major challenge to the UVH' prediction, the last effect was greatest for stimuli located at the point along the DHL at which the uncanny effect is predicted to occur. Taken together, the current data suggest that the UVH is most likely wrong in a number of points and essentially in its main prediction.

Implications and directions for future studies

The UVH can be understood in its present form as representing the initial stage of scientific enquiry (Northrop, 1966). In part because of this, empirical support for the uncanny effect has been inconsistent (e.g., Hanson, 2006; MacDorman, 2006; Tinwell and Grimshaw, 2009; Yamada et al., 2013; MacDorman et al., 2013). This has lead some researchers to question how uncanny-related research should now best proceed (e.g., Zlotowski et al., 2013). One approach is to understand Mori's visual hyperbole as serving as a useful heuristic device, rather than the object of research itself (as has been the case in most studies to date). As a hyperbole, this device makes a seemingly simply idea more easily accessible to exploratory thought in the face of incomplete and uncertain information about the potential problem that the UVH addresses. Exploratory thought requires the use of background knowledge from psychology and other domains of science in order to introduce to the UV problem known and communicable terminology, concepts, and methods, and for hypothesis testing a means to formulating clear definition of constructs, prediction of interrelationships, selection of experimental operations needed for testing, and provision of a framework for understanding data and a theoretical context for interpreting the findings (Fleck, 1979; Hempel, 1945; Northrop, 1966; Poundstone, 1988).. For example, theory and knowledge on perceptual and category processing,

perceptual decision making and CP can be used, as in the studies of this thesis, to develop, specify and generate testable ideas, for example, in relation to perceptual discrimination (Cheetham et al., 2011), or, for example, in relation to effects of high and low visual spatial frequency (Cheetham et al., 2013), effects of exposure duration and asymmetry in perceptual and category knowledge and experience, personality traits, priming (Cheetham et al., 2011; Saygin et al., 2011), or the true relevance of the *Hedonic Fluency Model* (Winkielman et al., 2003) for understanding uncanny-related effects.

The use of our *ad hoc* developed familiarity scale for assessing the UVH dimension of familiarity in Study C served the purpose of examining the conjectures of the UVH in terms that approximate most closely to the essence of the UVH itself. But, clearly, the UVH' ambiguous definition of the concept shinwakan and researchers' efforts to explicate what shinwakan might mean need to be superseded by the application of psychologically well-validated components of affective experience. In relation to the UVH, this application might entail considering affective experience along the DHL in terms of the constructs *valence* and *arousal* (e.g., Lang, et al., 1997). These constructs are intrinsic to all of Mori's illustrative examples of the uncanny effect (see Davitz, 1969), to the affective dimensions (e.g. likeability, pleasantness, familiarity) so often used to examine the uncanny effect (e.g., Green et al., 2008; Seyama and Nagayama, 2009, 2007; Dill et al., 2012; MacDorman and Ishiguro, 2006; Tinwell and Grimshaw 2009; Bartneck et al., 2007; Tinwell et al., 2011; Burleigh et al., 2013; MacDorman et al., 2013), and to all of the early theoretical accounts of uncanny experience (e.g. MacDorman, 2005).

7. References

- Albrecht, S., & Carbon, C. (2014) The Fluency Amplification Model: Fluent stimuli show more intense but not evidently more positive evaluations. *Acta Psychologica*, 148, 195-203.
- Balázs, B. Zur Kunstphilosophie des Films (1938). In: F.-J. Albersmeier (Hrsg.): *Theorie des Films*. Reclam, Stuttgart 1995, ISBN 3-15-009943-9, S. 204–226,
- Bartneck, C., Kanda, T., Ishiguro, H., & Hagita, N. (2007). "Is the Uncanny Valley an Uncanny Cliff?," in *Proceedings of the 16th IEEE, RO-MAN, Jeju*, 368–373.

- Beale, J. M., Keil, F. C. (1995). Categorical effects in the perception of faces. *Cognition*, 57, 217-239
- Burleigh, T. J., Schoenherr, J. R., & Lacroix, G. L. (2013). Does the uncanny valley exist? An empirical test of the relationship between eeriness and the human likeness of digitally created faces. *Computers in Human Behavior*. <http://dx.doi.org/10.1016/j.chb.2012.11.021>
- Campbell, R., Pascalis, O., Coleman, M., Wallace, S. B., & Benson, P. J. (1997). Are faces of different species perceived categorically by human observers? *Proc Biol Sci* 264, 1429-1434
- Cheetham, M., & Jancke, L. (2013). Perceptual and category processing of the Uncanny Valley Hypothesis' dimension of human likeness: some methodological issues. *J. Vis. Exp.* (76), e4375, doi:10.3791/4375
- Cheetham, M., & Jancke, L. (2013). Perceptual and category processing of the Uncanny Valley Hypothesis' dimension of human likeness: some methodological issues. *J. Vis. Exp.* (76).
- Cheetham, M., Suter, P., & Jancke, L. (2011). The human likeness dimension of the "uncanny valley hypothesis": behavioral and functional MRI findings. *Front Hum Neurosci.*, 5, 126.
- Cheetham, M., Pavlovic, I., Jordan, N., Suter, P., & Jancke, L. (2013). Category Processing and the human likeness dimension of the Uncanny Valley Hypothesis: Eye-Tracking Data. *Front Psychol.*, 4, 108. doi: 10.3389/fpsyg.2013.00108
- Cloutier J, Turk DJ, & Macrae CN. (2008). Extracting variant and invariant information from faces: the neural substrates of gaze detection and sex categorization. *Soc Neurosci.*, 3 (1):69-78.
- Cohen, B. (1989), *Developing Sociological Knowledge: Theory and Method*, Belmont, CA, Wadsworth Publishing.
- Davitz, J. R. (1969). *The language of emotion*. New York: Academic Press
- Decety, J. & Jackson, P. L. (2004). The functional architecture of human empathy. *Behav Cogn Neurosci Rev*, 3(2), 71-100.
- Dill, V., Flach, L., Hocevar, R., Lykawka, C., Musse, S., & Pinho, M. (2012). "Evaluation of the uncanny valley in CG characters," in *Intelligent Virtual Agents Lecture Notes in Computer Science*, eds N. Yukiko, M. Neff, A. Paiva, and M. Walker (Berlin: Springer), 511–513.
- Epley, N., Akalis, S., Waytz, A., & Cacioppo, J.T. (2008) Creating social connection through inferential reproduction. Loneliness and perceived agency in gadgets, gods, and greyhounds. *Psych Sci* 19:114-120.
- Etcoff, N. L., & Magee, J. J. (1992). Categorical perception of facial expressions. *Cognition*, 44, 227-240
- Fahrenberg, J., Myrtek, M., Pawlik, K., & Perrez, M. (2007). Ambulatory assessment - Monitoring behavior in daily life settings. A behavioral-scientific challenge for psychology. *European Journal of Psychological Assessment*, 23, 4, p.206-213.
- Fan, Y., Duncan, N. W., de Greck, M., & Northoff, G. (2011). Is there a core neural network in empathy? An fMRI based quantitative meta-analysis.. *Neurosci Biobehav Rev.*, 35 (3):903-11.
- Geller, T. (2008). Overcoming the uncanny valley. *IEEE Comput. Graph. Appl.* 28, 4, 11–17.

- Geven, A., Schrammel, J., & Tscheligi, M. (2006) "Interacting with embodied agents that can see: how vision-enabled agents can assist in spatial tasks," in *Proc. NordiCHI '06: Proceedings of the 4th Nordic conference on Human-computer interaction*, New York, NY, USA, 2006, pp. 135— 144.
- Gioia, D.A., & Pitre, E. (1990). *The Academy of Management Review*, 15, 4, pp. 584-602.
- Gray, H.M., Gray, K., & Wegner, D.M. (2007). Dimensions of mind perception. *Science*, 315 (5812), 619.
- Gray, J. A., & McNaughton, N. (2000). *The Neuropsychology of Anxiety*. London, Oxford University Press.
- Gray, K., & Wegner, D. M. (2012). Feeling robots and human zombies: Mind perception and the uncanny valley. *Cognition* 125:125-130.
- Green, R. D., MacDorman, K. F., Ho, C. -C., & Vasudevan, S. K. (2008). Sensitivity to the proportions of faces that vary in human likeness. *Computers in Human Behavior*, 24(5), 2456-2474
- Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: neural models of stimulus-specific effects. *Trends. Cogn. Sci.*, 10, 14-23
- Hanson, D (2006). Exploring the aesthetic range for humanoid robots. *Paper presented at the CogSci Workshop: Toward Social Mechanisms of Android Science*.
- Hanson, D., Olney, A., Prillman, S., Mathews, E., Zielke, M., Hammons, D., Fernandez, R., & Stephanou, H. E. (2005). Upending the Uncanny Valley. *20th National Conf. on Artificial Intelligence and the 17th Innovative Applications of Artificial Intelligence Conf* 1728-1729
- Harnad, S. R. (1987). Introduction: Psychological and cognitive aspects of categorical perception: A critical overview. In *Categorical perception: The groundwork of cognition*, Harnad, S. R., ed. (New York: Cambridge University Press. Haruno & Kawato), pp. 1-25
- Harnad, S. R. (1987). Introduction: Psychological and cognitive aspects of categorical perception: A critical overview. In H. SR (Ed.), *Categorical perception: The groundwork of cognition* (pp. 1-25). New York: Cambridge University Press.
- Hegel, F., Eyssel, F., & Wrede, B. (2010). The Social Robot Flobi: Key Concepts of Industrial Design“. In *Proceedings of the International Symposium on Robot and Human Interactive Communication (RO-MAN)*, Viareggio (Italy), 107-112.
- Hegel, F., Spexard, T., Horstmann, G., Vogt, T., & Wrede, B. (2006). Playing a different imitation game: Interaction with an empathic android robot. In *Proceedings of the IEEE-RAS International Conference on Humanoid Robots (Humanoids 06)*, Genova, Italy. 56-61.
- Hempel, C. G. (1945). Studies in the logic of confirmation. *Mind*, 54, 1–26, 97–121.
- Ho, C. C., MacDorman, K. F., & Pramono, Z. A. D. (2008). Human emotion and the uncanny valley: A GLM, MDS, and isomap analysis of robot video ratings. In *ACM/IEEE international conference on human–robot interaction*. 169-176.
- Horton, D., & Wohl, R. (1956). Mass communication and para-social interaction: Observation on intimacy at a distance. *Psychiatry*, 19, 215–229.

- Jiang, X., Rosen, E., Zeffiro, T., Vanmeter, J., Blanz, V., & Riesenhuber, M. (2006). Evaluation of a shape-based model of human face discrimination using fMRI and behavioral techniques. *Neuron*, 50, 159-172
- Karnadewi F, & Lipp OV. (2011). The processing of invariant and variant face cues in the Garner Paradigm. *Emotion*, 11(3):563-71.
- Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1997). Motivated attention: Affect, activation, and action. In P. J. Lang, R. F. Simons, & M. T. Balaban (Eds.), *Attention and orienting: Sensory and motivational processes* (pp. 97-135). Mahwah, NJ: Erlbaum.
- Lieberman, A.M., Hariris, K.S., Hoffman, H.S., & Griffith, B.C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *J. Exp. Psychol.* 54, 358-368
- Lienert, G. A. & Raatz, U. (1994). Testaufbau und Testanalyse (5., völl. Neubearb. u. erw. Aufl.). Weinheim: Psychologie Verlags Union.
- Livingston, K. R., Andrews, J. K., & Harnad, S. (1998). Categorical perception effects induced by category learning. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 24(3), 732– 753.
- Lombard, M. & Ditton, T. (1997). At the Heart of It All: The Concept of Presence. *Journal of Computer-Mediated Communication*. *Volt*, 3(2).
- Looser, C.E. & Wheatley, T. (2010). The tipping point of animacy: How, when, and where we perceive life in a face. *Psychological Science*. 21, 1854-1862
- Looser, C.E., & Wheatley, T. (2010). The tipping point of animacy: How, when, and where we perceive life in a face. *Psychological Science*, 21, 1854–1862
- Lynham, S. A. (2002). The general method of theory-building research in applied disciplines. *Advances in Developing Human Resources*, 4 (3), 221-235.
- Maccoby, E.E., Wilson, W.C., 1957. Identification and observational learning from films. *Journal of abnormal Social Psychology* 55, 76–87.
- MacDorman, K. F., & Ishiguro, H. (2006). "The uncanny advantage of using androids in cognitive and social science research". *Interaction Studies*, 7, 297-337
- MacDorman, K.F., Srinivas, P., & Patel, H., (2013). The uncanny valley does not interfere with level 1 visual perspective taking, *Computers in Human Behavior*, 29, 4, 1671-1685
- MacGillivray, C. (2007). How psychophysical perception of motion and image relates to animation practice. *Imaging and Visualization*, International Conference on Computer Graphics, 81–88.
- Maddock, S. Edge J., & Sanchez M (2005). "Movement realism in computer facial animation," in Workshop on Human-Animated Characters Interaction at *The 19th British HCI Group Annual Conference HCI 2005: The Bigger Picture*, Edinburgh.
- Minato, T., Shimada, M., Ishiguro, H., & Itakura, S. (2004). Development of an android robot for studying human-robot interaction. *Innovations in applied artificial intelligence*, 424-434.
- Mori, M (1970). Bukimi no tani "the uncanny valley" (k. f. MacDorman & t. Minato, trans.). *Energy*, 7(4):33:35, 1970.
- Nothrop, F. S. C., (1967). *The Logic of the Sciences and the Humanities*. New York, Meridian Books, World Publishing Company.

- Pastore, R. E. (1987). Categorical Perception: Some Psychophysical Models. In S. Harnad (Ed.), *Categorical perception: The groundwork of Cognition*. New York: Cambridge University Press.
- Popper, K. (2004). *Conjectures and refutations: the growth of scientific knowledge*. London: Routledge.
- Poundstone, W. (1988). *Labyrinths of Reason. Puzzles, Paradoxes and the Frailty of Knowledge*. New York: Anchor Press/Doubleday.
- Power, G., Wills, G. and Hall, W., (2002). User Perception of Anthropomorphic Characters with Varying Levels of Interaction. *Proc. HCI '02 UK*, 37- 51.
- Repp, B.H. (1984).Categorical perception: Issues, methods, findings. *Speech Lang. Adv. Basic Res. Pract.* 10, 243-335
- Roberts, R., and Kreuz, R. (1994). Why do people use figurative language. *Psychological Science*, 5, 159-163.
- Tinwell A, Grimshaw, M, & Williams A, (2010). Uncanny behaviour in survival horror games. *Journal of Gaming & Virtual Worlds*, 2(1), 3-25.
- Winkielman P, Schwarz N, Fazendeiro T, Reber R (2003) The hedonic marking of processing fluency: Implications for evaluative judgment. In: Klauer JMKC, editor. *The psychology of evaluation: Affective processes in cognition and emotion*. Mahwah, NJ: Lawrence Erlbaum. pp. 189–217.
- Yamada, Y., Kawabe, T., & Ihaya, K. (2013). Categorization difficulty is associated with negative evaluation in the “uncanny valley” phenomenon. *Japanese Psychological Research*, 55 (1), 20–32 doi.org/10.1111/j.1468-5884.2012.00538.x.
- Zlotowski, J., Proudfoot, D., & Bartneck, C. (2013). More Human Than Human: Does The Uncanny Curve Really Matter? *Proceedings of the HRI2013 Workshop on Design of Humanlikeness in HRI from uncanny valley to minimal design*. Tokyo, pp. 7-13
- Schindler K, Van Gool L, & de Gelder B.(2008). Recognizing emotions expressed by body pose: a biologically inspired neural model. *Neural Netw.*,21(9):1238-46
- Schneider, E., Wang, Y., & Yang, S. (2007). Exploring the uncanny valley with japanese video game characters. In *Situated Play*, B. Akira, Ed. 546–549.
- Seyama, J., & Nagayama, R. S. (2007).The Uncanny Valley: Effect of Realism on the Impression of Artificial Human Faces. *Presence: Teleoperators and Virtual Environments*, 16, 337-351
- Seyama, J., and Nagayama, R. S. (2007). The Uncanny Valley: Effect of Realism on the Impression of Artificial Human Faces. *Presence: Teleoperators and Virtual Environments*, 16, 337-351
- Tellegen, A., & Atkinson, G. (1974). Openness to absorbing and self-altering experiences ("absorption"), a trait related to hypnotic susceptibility. *Journal of Abnormal Psychology*, 83, 268-277.
- Tinwell, A. (2009).Uncanny as usability obstacle. *HCI* 12, 622-631.
- Tinwell, A., Grimshaw, M., & Williams, A. (2011). The Uncanny Wall. *International Journal of Arts and Technology* 4, 326-341.
- Van de Ven, A.H. (1989). Nothing is quite so practical as a good theory. *The Academy of Management Review*,14(4), 486-489.

- Waytz, A., & Norton, M. (2014). Botsourcing and outsourcing: Robot, British, Chinese, and German workers are for thinking--not feeling--jobs. *Emotion, 14*(2):434-44.
- Waytz, A., Gray, K., Epley, N., & Wegner, D., M. (2010). Causes and consequences of mind perception. *Trends Cogn Sci. 14*(8):383-8. doi: 10.1016/j.tics.2010.05.006.

8. Curriculum Vitae

Name: Marcus Cheetham

Personal: Born 08.07.1966 in Bowden, England

Address: University of Zurich
Institute of Psychology
Division Neuropsychology
Binzmühlestrasse 14/25
CH-8050 Zürich
Switzerland

Telephone: 0041 (0)44 635 7393

E-mail: m.cheetham@psychologie.uzh.ch

Academic Education and Qualifications:

01.05.2013 - to date Senior Research Assistant (Oberassistent ad personam)
Dept. of Neuropsychology, Institute of Psychology,
University of Zurich, Switzerland

01.11.2011 - 30.04.2013 Postdoctoral researcher
Division Neuropsychology, Institute of Psychology,
University of Zurich, Switzerland

01.03. 2007 - 31.10.2011 Research Assistant / Ph.D. student
Division Neuropsychology, Institute of Psychology,
University of Zurich, Switzerland

Ph.D. Thesis: " The Uncanny Valley Hypothesis:
behavioural, eye-movement, and functional MRI findings"
(summa cum laude).

Supervisors:

Lutz Jäncke, Prof. Dr.; Martin Meyer, Prof. Dr.

01.10.1995 - 30.09.2003 Master's Degree (Dipl.-Psych) in Psychology
Albert-Ludwigs University,
Freiburg i.Br., Federal Republic of Germany.

Master's Thesis: "Development and validation of a self-report instrument for eating behaviour according to the DSM-IV" (written in German)
Supervisor: Franz Caspar, Prof. Dr.

Research and Clinical Experience:

01.05.2013 - to date Senior Research Assistant (Oberassistent ad personam)
Dept. of Neuropsychology, Institute of Psychology,
University of Zurich, Switzerland

01.11.2011 - 30.04.2013 Postdoctoral researcher
Division Neuropsychology, Institute of Psychology,
University of Zurich, Switzerland

01.03. 2007 - 31.10.2011 Research Assistant / Ph.D. student
Division Neuropsychology, Institute of Psychology,
University of Zurich, Switzerland

01.01.2001 - 31.08.2001 Internship "Diagnostic clinical interviews and Eating Disorders". University of Basel, Switzerland
Supervisor: Dr. S. Munsch

01.09.2001 - 31.12.2001 Internship "Diagnostic clinical interviews and Eating Disorders". Cantonal Hospital, Basel, Switzerland
Supervisor: Dr. S. Munsch

- 01.07.2000 - 30.09.2000** Internship "Neuropsychological Diagnosis and treatment"
Albert-Ludwigs University,
Freiburg i.Br., Federal Republic of Germany.
Supervisor: G. Lucius-Hoene, Prof. Dr.
- 01.04.2000 - 31.07.2000** Tutorial Assistant "Elements of Neurobiology II"
Albert-Ludwigs University,
Freiburg i.Br., Federal Republic of Germany.
Supervisor: U. Halsband, Prof. Dr.
- 01.04.2000 - 31.07.2000** Internship "Techniques, data acquisition
and assessment in fMRI "
Neurozentrum, Albert-Ludwigs University,
Freiburg i.Br., Federal Republic of Germany.
Supervisor: M. Green, Prof. Dr.
- 01.04.2000 - 31.07.2000** Tutorial Assistant "Elements of Neurobiology I"
Albert-Ludwigs University,
Freiburg i.Br., Federal Republic of Germany.
Supervisor: M. Peper, Prof. Dr

Peer - Reviewed Publications:

Cheetham, M., Suter, P., & Jancke, L. (under review). Perceptual discrimination difficulty and familiarity in the Uncanny Valley.

Meier, L.M., De Matos, N., Bruegger, M., Lukic, N., Cheetham, M., Ettlin, D., Jancke, L., & Lutz, K. (2014). Equal pain – Unequal fear response: Enhanced susceptibility of tooth pain to fear conditioning. *Front. Hum. Neurosci.*, doi: 10.3389/fnhum.2014.00526.

Cheetham, M., Hänggi, J., and Jancke, L (2014). Identifying with fictive characters: structural brain correlates of the personality trait 'fantasy'. Soc Cogn Affect Neurosci., doi: 10.1093/scan/nst179.

Lutz, K., Pourger, R., Cheetham, M., and Jancke, L (2013). Development of ERN together with an internal model of audio-motor associations. Front Hum Neurosci., doi: 10.3389/fnhum.2013.00471.

Cheetham, M., Pavlovic, I., Jordan, N., Suter, P., Jancke, L. (2013). Perceptual decision-making, the dimension of human likeness, and the uncanny valley hypothesis: An eye movement study. Frontiers in Psychology. doi: 10.3389/fpsyg.2013.00108.

Cheetham, M., and Jancke, L. (2013). Perceptual and category processing of the Uncanny Valley Hypothesis' dimension of human likeness: some methodological issues. J. Vis. Exp. (76). doi: 10.3791/4375.

Havranek, M., Langer, N., Cheetham, M., & Jancke, L. (2012). Perspective and agency during video gaming influences spatial presence experience and brain activation patterns. Behav Brain Funct. 8, (34). doi: 10.1186/1744-9081-8-34.

Cheetham, J.M., Rahm, B., Kaller C.P., & Unterrainer J.M., (2012). Visuospatial over verbal demands in predicting Tower of London planning tasks. British Journal of Psychology, 103 (1), 98–116. doi: 10.1111/j.2044-8295.2011.02049.x

Cheetham, M., Suter, P., Jancke, L. (2011) The human likeness dimension of the "uncanny valley hypothesis": behavioral and functional MRI findings. Front Hum Neurosci.;5:126. doi: 10.3389/fnhum.2011.00126.

Cheetham, M., Pedroni, A.F., Antley, A., Slater, M., Jancke, L. (2009) Virtual milgram: empathic concern or personal distress? Evidence from functional MRI and dispositional measures. Front Hum Neurosci.;3:29 doi: 10.3389/neuro.09.029.2009

Jancke, L., Cheetham, M., & Baumgartner, T. (2009). Virtual reality and the role of the prefrontal cortex in adults and children. *Front Hum Neurosci* .3(1):52-9. doi: 10.3389/neuro.01.006.2009

Fahrenberg, J. & Cheetham, M. (2007). Assumptions About Human Nature and the Impact of Philosophical Concepts on Professional Issues: A Questionnaire-Based Study with 800 Students From Psychology, Philosophy, and Science. *Philosophy, Psychiatry, and Psychology* 14 (3):pp. 183-201.

Fahrenberg, J. & Cheetham, M. (2000). The Mind-Body Problem as Seen by students of Different Disciplines. *Journal of Consciousness Studies* 7 (5):47-59.

Ad Hoc Reviewer

Advances in Cognitive Psychology

European Journal of Social Psychology

International Journal of Social Robotics

Frontiers in Cognitive Science

PLOS ONE

Editor

Frontiers Research Topic in Frontiers of Neuroscience:
The Uncanny Valley and beyond.

<http://www.frontiersin.org>